Characterization of the KASH domain gene *unc-83* and the pseudogene *F55A3.7*

DISSERTATION

zur Erlangung des akademischen Grades

Doctor rerum naturalium

(Dr. rer. nat.)

eingereicht an der

Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

von

Mag. rer. nat. Andreas Ofenbauer

Präsidentin der Humboldt-Universität zu Berlin

Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

Prof. Dr. Bernhard Grimm

Gutachter/in: 1. Prof. Thomas Sommer

2. Dr. Baris Tursun

3. Prof. Andrew Plested

Datum der Einreichung: 22.11.2018

Datum der Promotion: 24.05.2019

Zusammenfassung

Mein Ziel war es, genetische Faktoren in *C. elegans* zu identifizieren, die eine Rolle bei induzierter Transdifferenzierung durch Missexpression des Transkriptionsfaktors (TF) HLH-1, welcher das Wurmhomolog des myogenen bHLH TF MyoD ist, spielen. Ich entwickelte hierzu einen semiautomatischen Hochdurchsatz-Vorwärtsgenetik-Screen, indem ich EMS Mutagenese mit dem Biosorter-System (Union Biometrica) kombinierte. Missexprimiert kann HLH-1 Muskelentwicklung in frühen embryonalen Zellen induzieren, aber terminal differenzierte Zellen in älteren Tieren sind gegenüber HLH-1-induziertem direktem Reprogrammieren resistent. Um Mechanismen zu identifizieren die HLH-1-induziertem direktem Reprogrammieren in terminal differenzierten Zellen entgegenwirken, nutzte ich für meinen Screen eine transgene Linie, die ektopische Expression von *hlh-1* ermöglicht und des Weiteren einen transkriptionellen Reporter für Muskelzellen besitzt (*myo-3p::gfp::NLS*). Mit diesem Ansatz ist es mir gelungen, die Mutante *bar18* zu isolieren, die eine Anhäufung an Muskelzellkernen um den posterioren Teil des Pharynx zeigt. Ich identifizierte den mutierten Lokus, indem ich das gesamte Genom sequenzierte und charakterisierte den mutanten Phänotyp im Detail.

Zusätzlich war ich bei der Charakterisierung von Faktoren, die das Umprogrammieren zu neuronalen Zellen in C. elegans verhindern, beteiligt. Dabei stand der sogenannte FACT-Komplex im Focus, welcher mittels eines genom-weiten RNAi-Screen in unserer Arbeitsgruppe identifiziert wurde[1]. Dieser Rückwärtsgenetik-Screen zielte, im Gegensatz zum Vorwärtsgenetik-Screen den ich durchführte, darauf ab, Faktoren zu identifizieren, die bei der induzierten Transdifferenzierung durch Missexpression von dem TF CHE-1 eine Rolle spielen. CHE-1 ist ein Zinkfinger-TF und essentiell für die terminale Differenzierung der glutamatergen ASE-Neuronen. Interessanterweise ist eine der FACT-Komplex-Untereinheiten, *spt-16*, das parentale Gen zu dem bislang nicht charakterisierten Pseudogen *F55A3.7*. Eine putative Null-Mutante von *F55A3.7*, in Kombination mit ubiquitärer Überexpression von CHE-1, zeigte einen Keimzellen-zu-Neuronen Transdifferenzierungsphänotyp ähnlich dem Phänotypen, der nach dem Knock-down der FACT-Komponente *hmg-3* beobachtet wird. Unseres Wissens nach ist dies das erste Beispiel einen Pseudogens, dessen Knock-down dazu führt, dass ein bestimmtes Gewebe durch einen terminalen Selektor-TF reprogrammiert werden kann, dessen Expression unter normalen Konditionen dies nicht zur Folge hätte. Aufgrund dieser Einzigartigkeit, habe ich das Pseudogen *F55A3.7* charakterisiert und außerdem versucht, einen Mechanismus zu finden, wie *F55A3.7* die Keimbahnidentität schützt.

Abstract

My aim was to identify and characterize genetic factors in *C. elegans* that play a role in induced transdifferentiation by mis-expressing the transcription factor (TF) HLH-1, which is the worm homolog of the myogenic bHLH TF MyoD. For this, I developed a semi-automated high-throughput forward genetic screen combining EMS mutagenesis with the Biosorter system (Union Biometrica). When mis-expressed, HLH-1 induces the muscle fate in early embryonic cells, but terminally differentiated cells in older animals are resistant to HLH-1-induced direct reprogramming. In order to identify mechanisms that antagonize HLH-1-induced reprogramming in ectopic tissues, I used a transgenic line allowing ectopic expression of the *hlh-1* gene in combination with a reporter for muscle fate (*myo-3p::gfp::NLS*) for my screen. Using this approach, I isolated the mutant *bar18*, showing an accumulation of muscle cell nuclei around the posterior pharyngeal bulb of the worm. I identified the mutated locus using whole genome sequencing (WGS) and characterized the identified gene and the mutant phenotype further.

Additionally, I was also involved in helping to characterize the FACT complex, which was identified through a whole-genome RNAi screen conducted by my colleague Ena Kolundžić[1]. This reverse genetic screen, in contrast to the forward genetic screen that I performed, aimed at identifying factors that play a role in induced transdifferentiation by mis-expressing the TF CHE-1, a Zn-finger TF essential for terminal differentiation of glutamatergic ASE neurons. Interestingly, one of the FACT complex members, *spt-16*, is the parental gene of a previously uncharacterized pseudogene named *F55A3.7*. A putative null mutant of *F55A3.7,* combined with broad overexpression of CHE-1, showed a germ cells to neurons transdifferentiation phenotype similar to the phenotype observed after depleting the FACT member *hmg-3*. To our knowledge, this is the first example of a pseudogene whose depletion leads to the permissiveness of a certain tissue to be reprogrammed when challenged by a terminal selector TF. Due to this uniqueness, I characterized the pseudogene *F55A3.7* and tried to find a potential mechanism for how *F55A3.7* safeguards germline identity.

Keywords:

LINC, FACT, *unc-83*, *F55A3.7*

Table of Content

Zusammenfassung **2**

Abstract **3**

Dedication **7**

List of Abbrevations **8**

Preface **10**

1 Introduction **11**

1.1 Maintenance of cell identities and cell function 11

1.2 *C. elegans* as a model system 12

1.2.1 Eutely and cell lineage in *C. elegans* 16

1.2.2 The *C. elegans* muscle system 17

1.2.3 The *C. elegans* gonad 20

1.3 Principles of genetic screens in *C. elegans* 23

1.3.1 Forward genetic screens 23

1.3.2 Reverse genetic screens 24

1.4 Forward genetic screens targeted at isolating cell lineage mutants identified members of LINC complexes in *C. elegans* 31

1.4.1 LINC complexes 31

1.4.2 The UNC-83/UNC-84 LINC complex 32

1.5 Reverse genetics identified reprogramming barriers in *C. elegans* 34

1.5.1 FACT (FAcilitates Chromatin Transcription) complex 35

1.6 Pseudogenes 36

1.6.1 *F55A3.7* is a pseudogene related to *spt-16* 39

1.7 Aim of this thesis 41

2 Results **42**

2.1 A forward genetic screen reveals the requirement of the LINC member UNC-83 during positioning of body wall muscle nuclei in *C. elegans* 42

2.1.1 Additional muscle nuclei around the posterior pharyngeal bulb in *bar18* mutant animals are independent of ectopic *hlh-1* induction 42

2.1.2 Whole-genome sequencing in conjunction with a SNP Mapping Strategy revealed that the mutated locus belongs to *unc-83* 43

2.1.3 The accumulation of muscle nuclei around the posterior pharyngeal bulb in *unc-83(bar18)* mutant animals is a cell autonomous effect 45

2.1.4 Accumulated nuclei around the posterior pharyngeal bulb in *unc-83(bar18)* animals are not due to an overall increase in the amount of muscle cells 46
2.1.5 The UNC-83/UNC-84 LINC complex is required for proper body wall muscle nuclei positioning in *C. elegans* 47
2.1.6 Mutations of the LINC members *unc-83* and *unc-84* cause motility defects in *C. elegans* 48
2.1.7 Lack of the LINC member UNC-83 shortens the life span of *C. elegans* 51
2.1.8 A HLH-1::tagRFP fusion protein is actively degraded by the proteasome in *C. elegans* 54
2.2 A reverse genetic screen identifies members of FACT and the pseudogene *F55A3.7* as barriers of induced germ cell to neuron conversion in *C. elegans* 57
2.2.1 The pseudogene *F55A3.7* is being transcribed and spliced 58
2.2.2 *F55A3.7* RNA is not translated into a protein 59
2.2.3 *F55A3.7* ncRNA does not affect mRNA and protein levels of its mother gene *spt-16*, nor those of *hmg-3* or *lin-53* 60
2.2.4 *F55A3.7* ncRNA acts in trans 61
2.2.5 Gonad specific RNA-seq and ATAC-seq reveal many up- and downregulated loci 62
3 Discussion 68
3.1 Forward vs. reverse genetic screens 68
3.2 A forward genetic screen aiming to find reprogramming barriers in *C. elegans* reveals something unexpected 71
3.3 The role of the UNC-83/UNC-84 LINC complex in *C. elegans* body wall muscle cells 72
3.4 The ncRNA *F55A3.7* does not affect its mother gene *spt-16* and acts in trans to protect germline fate 74
3.5 Gonad specific transcriptome and epigenetic landscape analysis in *F55A3.7* depleted worms 75
4 Material & Methods 77
4.1 Nematode culture 77
4.2 Transgenesis in *C. elegans* 77
4.2.1 Microinjection into *C. elegans'* gonads 77
4.2.2 Integration of extrachromosomal arrays with γ-irradiation 77
4.2.3 Universal MosSCI 78
4.2.4 Gene tagging using CRISPR/Cas9 78
4.3 Transgenic crosses of *C. elegans* 78

4.4 Identification of the mutated *bar18* locus with WGS 79
4.5 Microscopic analysis of *bar18* phenotype 79
4.6 Lifespan assays 80
4.7 Swimming assays 80
4.8 Exploration assays 80
4.9 RNAi experiments 81
4.10 Immunostaining of *C. elegans* 81
4.11 cDNA preparation and qRT-PCR 82
4.12 Western blots 82
4.13 Transdifferentiation assays (heat-shock induction of *che-1*) 83
4.14 RNA-seq 83
4.15 ATAC-seq 83
4.16 List of worm strains used 84
4.17 List of plasmids used 85
5 References 85
Publications 101
Acknowledgements 102
Declaration 103

Dedication

To the ones that always loved and supported me:

Mama, Papa, David und Tante Christa.

List of Abbrevations

AB	Antibody
ATAC-seq	Assay for Transposase Accessible Chromatin with high-throughput sequencing
cDNA	Complementary DNA
Cas9	A nuclease from *Streptococcus pyogenes*
Cas9n	Cas9 nickase
DSB	Double-strand break
DNA	Deoxyribonucleic acid
EMS	Ethyl-Methane-Sulfonate
ESC	Embryonic stem cell
EV	Empty vector
esiRNA	Endoribonuclease-prepared siRNA
fCas9	Catalytically inactive Cas9 fused with FokI domain
FokI	A restriction enzyme of *Flavobacterium okeanokoites*
FUDR	5-Fluoro-2′-deoxyuridine
GFP	Green fluorescent protein
GULO	L-gulono-γ-lactone oxidase
HDR	Homology-directed repair
HR	Homologous recombination
iPSC	induced Pluripotent stem cell

NGS	Next generation sequencing
NLS	Nuclear localization signal
ORF	Open reading frame
PAM	Protospacer-Adjacent Motif
PCA	Principle component analysis
RNA	Ribonucleic acid
RNAi	RNA interference
RNA-seq	RNA sequencing
SD	Standard deviation
SED	Standard error of the mean
siRNA	Small interfering RNA
TALEN	Transcription activator-like effector nuclease
TF	Transcription factor
UTR	Untranslated region
WGS	Whole-genome sequencing
WT	Wild-type
ZFN	Zinc finger nuclease

Preface

The basis for this research, for my thesis, originally stemmed from my passion for genetics and developmental biology. I was always curious about the act of undifferentiated cells undergoing specialization, forming tissues, organotypic structures and whole organisms. The translation of intracellular alterations into observable tissue patterns with their associated functions is extremely exciting and lets us embrace the beauty of nature on a molecular level.

My ride was not easy though and sometimes I had the feeling there were more downs than ups. For three years I was riding a dead horse - a crazy project that I had to drop because it was leading nowhere. On the other hand, isn't a smooth ride boring? And whose PhD is running smoothly anyways? The ones I've heard of are fairytales I believe.

What really matters is the knowledge I gathered, the skills I learned, the amazing experiences I had and the wonderful people I met on my way. I would have never wanted to miss out of all these opportunities and the most inspiring and interesting time of my life so far.

1 Introduction

1.1 Maintenance of cell identities and cell function

The idea that terminally differentiated cells are irreversible committed to a certain cell identity has been challenged several times throughout the history of modern biology. Pioneering work that showed that terminally differentiated cells still possess a certain plasticity has been done by Gurdon in the fifties[2] and Wilmut in the nineties[3]. Both could show that the transplantation of a somatic, terminally differentiated nucleus into an unfertilized denucleated oocyte could eventually lead to reprogramming of the somatic nucleus and ultimately resulted in the cloning of frogs[2] and the sheep Dolly[3]. Direct reprogramming or transdifferentiation by ectopic expression of cell-fate inducing TFs was first demonstrated in the eighties by the Weintraub lab[4]. Davis and colleagues could show that transfecting fibroblasts with a cDNA encoding the TF MyoD is sufficient to convert them into contracting myoblasts. Another milestone in the field was the discovery by Takahashi and Yamanaka, that four specific transcription factors (TFs), OCT4, SOX2, KLF4 and c-MYC, also collectively referred to as OSKM, could convert mouse fibroblasts to a pluripotent cell type called induced pluripotent stem cells (iPSCs)[5]. iPSCs have similar morphology, gene expression profiles and growth properties as embryonic stem cells (ESCs). Importantly, those cells form teratomas when injected into mice, which are tumors that consist of cells from all three germ layers, thus ultimately confirming their pluripotency.

The concept that terminally differentiated cells are stably, rather than terminally, committed[6], has inspired many labs to screen for reprogramming barriers in different species. Reprogramming barriers usually ensure that specific cell identities are being maintained in an organism. *C. elegans* is a powerful model organism to identify and study such reprogramming barriers *in vivo*. The chromatin regulator FACT (facilitates chromatin transcription) for instance, was identified in *C. elegans* as a reprogramming barrier for the transdifferentiation of germ and intestinal cells to neurons[1]. Strikingly, FACT's role as a barrier to cell fate conversion is conserved in *H. sapiens*, confirming the potential of using *C. elegans* as a system to identify evolutionary conserved reprogramming barriers.

Besides reprogramming barriers, there are many other maintaining mechanisms and factors ensuring that a certain cell type is properly fulfilling its specific functions. LINC (Linker of Nucleoskeleton and Cytoskeleton) complexes, for instance, fulfill many different roles that are crucial in this regard, from ensuring proper nuclear movement and anchoring, to moving meiotic chromosomes and telomeres or sensing mechanic stimuli[7–10]. They are widely conserved too, including nematodes, plants and mammals and a lot of our current understanding about LINC complexes is due to studies that have been done in the worm.

C. elegans with its invariant cell lineages is an excellent model to study maintenance in general, be it maintenance of cell identities, maintenance of cell number or cell position, amongst many others. The specific advantages of *C. elegans* as a model organism are being discussed in the next section.

1.2 *C. elegans* as a model system

Caenorhabditis elegans (*C. elegans*) is a non-parasitic nematode of about 1 mm in length and lives on rotten vegetable or fruit matter that contains a bacterial food source[11]. It can be found all over the world and was first isolated as a model organism in Bristol in 1963 by Sydney Brenner[12]. Under laboratory conditions, it is usually grown on agar plates seeded with *Escherichia coli* (*E. coli*) as a food source or in liquid culture containing *E. coli*.

C. elegans has a cylindrical body shape and consists, like other nematodes, of several specialized tissues (Figure 1). Cuticle, hypodermis, excretory system, neurons and body wall muscles form the outer tube of the worm, while pharynx, intestine and gonad belong to the inner tube. Both tubes are separated by the so-called pseudocoelomic space.

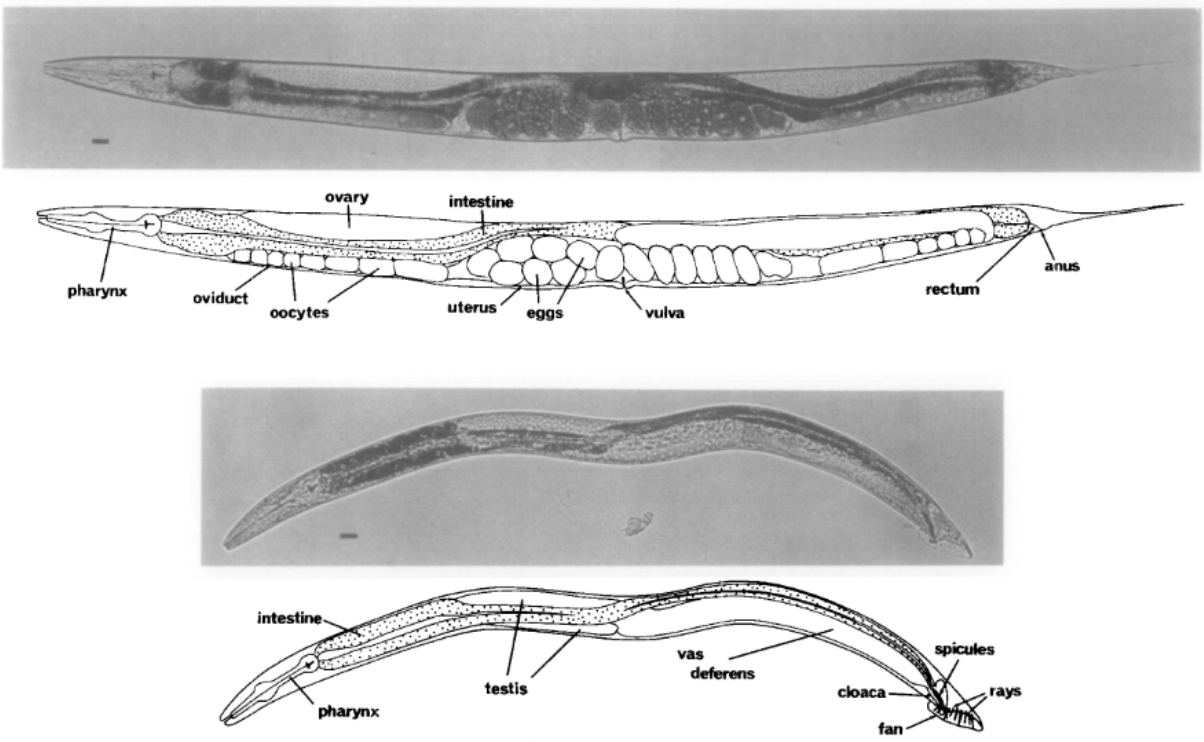


Figure 1. ***C. elegans.*** Adult hermaphrodite (above) and male (below). Lateral views, bright field illumination. Scale bar = 20µm. Figure adapted from Sulston and Horvitz, 1977[13].

They are translucent, which easily allows microscopic imaging of all tissues *in vivo*, using e.g. differential interference contrast (DIC) microscopy, but also fluorescent microscopy, which takes advantage of the possibility to express reporter genes in a cell- or tissue-specific manner (Figure 2).

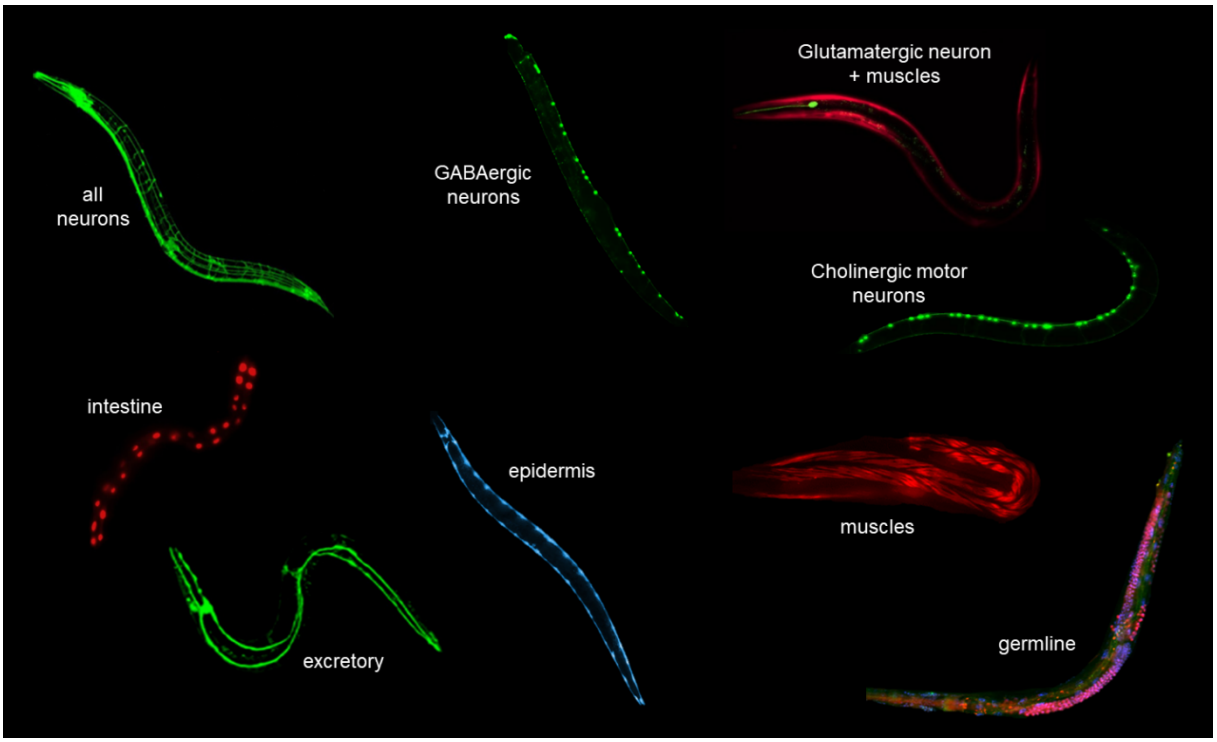


Figure 2. Tissue specific *in vivo* labelling in *C. elegans*. *C. elegans* is a translucent organism. Using tissue specific promoters to drive the expression of reporter genes allows labeling of subsets of cells in the worm. Picture: Tursun et al., Hobert et al.

Besides its translucency, *C. elegans* has several other features that make it an excellent model organism. It has a rapid life cycle (~3 days from egg to an egg-laying adult at 25 °C) and exists primarily as a self-fertilizing hermaphrodite, although males occur spontaneously with a frequency of < 0,2 % within a population. Furthermore, it is a eutelic organism (Section 1.1.1), meaning that it is composed of a fixed number of somatic cells: hermaphrodites consist of 959 cells, whereas males have 1.031 cells. Self-fertilized hermaphrodites lay up to 300 eggs, while hermaphrodites inseminated by males lay up to 1000.

The life cycle of *C. elegans* consists of embryogenesis (~13h), the larval stages L1-L4 (~28h) and adulthood (Figure 3). Embryogenesis can be divided into an *in utero* (mostly proliferation) and an *ex utero* development, which starts after egg laying coinciding roughly with the beginning of gastrulation[14,15]. The so-called 'lima bean' stage marks the beginning of morphogenesis and elongation. After a quickening phase, marked by the beginning of muscle movements, animals hatch and pass through the larval stages L1-L4. The reproductive system is already established during larval development and is fully developed as the worms reach adulthood. At 22-25 °C, hermaphrodites start laying their first eggs approximately 45-50 h after hatching. Unfavorable conditions such as starvation, high temperatures or crowding can

lead to alternative larval forms like L1 arrest or Dauer. During these stages, further development is arrested and locomotion is reduced. Once more advantageous conditions return, development continues: L1 arrested animals molt and develop to L2 larvae, while Dauer worms develop into L4 larvae.

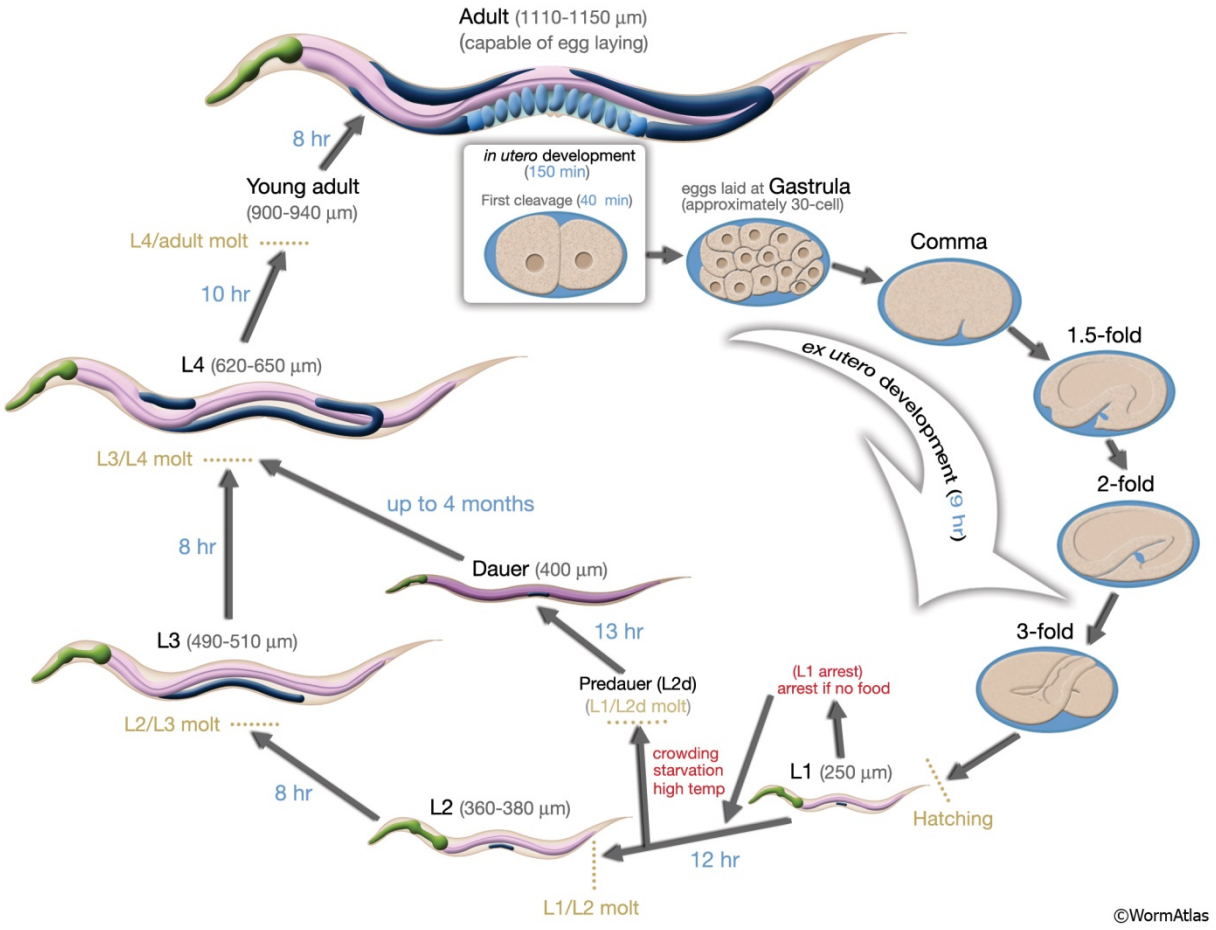


Figure 3. Life cycle of *C. elegans* at 22 °C. The life cycle of *C. elegans* consists of embryogenesis (~13h), the larval stages L1-L4 (~28h) and adulthood. Unfavorable conditions like starvation, high temperatures or crowding can lead to alternative larval forms like L1 arrest or Dauer. Picture taken from www.wormatlas.org

C. elegans was the first multi-cellular organism to have its genome sequenced (*C. elegans* Sequencing Consortium 1998). It has five autosomes and one sex chromosome: hermaphrodites carry a matched pair of sex chromosomes (XX), while males carry only one (X0)[16]. Although the number of protein coding genes between humans and *C. elegans* is comparable (both have around ~20000), the genome of *C. elegans* is much smaller as compared to humans (~100Mb organized in 6 chromosomes vs. ~3200Mb organized in 23 chromosomes). Remarkably, over two-thirds of human protein coding genes have homologs in *C. elegans*[17]. Important molecular pathways such as Wnt, Notch and Insulin signaling are conserved between *C. elegans* and humans[18].

Taken together, *C. elegans* offers several features that make it an excellent model organism: its small size and translucency, a rapid life cycle and being primarily hermaphroditic, a well annotated genome allowing genetic screens and manipulations as well as having homologs to many human protein coding genes, including many important molecular pathways. Furthermore, *C. elegans* offers eutely and its complete lineage has been mapped (Section 1.1.1).

1.2.1 Eutely and cell lineage in *C. elegans*

Eutely, or cell constancy, refers to organisms that have a constant number of somatic cells when they reach adulthood[19,20]. After fertilization and initial proliferation, each embryonic cell possesses a limit for total cell divisions. After each cell has reached its limit and the total number of cells is established, somatic development stops and the organism can only become bigger due to cell enlargement (hypertrophy) rather than through cell division.

Eutelic animals are usually microscopically small and consist of less than 100.000 cells. Examples include rotifers, tardigrades, gastrotrichs, dicyemids and nematodes. Amongst nematodes, maritime nematodes are an exception, since they don't display eutely and thus have an undefined number of total cells[21]. Eutely allows the analyzation of the developmental history of each cell in the whole organism, tracing its individual lineage back to fertilization. It is also a nicely exploitable trait for experiments that aim to identify genes involved in lineage decisions and (terminal) differentiation. This is especially true for forced transdifferentiation experiments, as the readout of such experiments is usually a tissue-specific reporter gene.

C. elegans is the best studied eutelic organism to date and its whole cell lineage is mapped[13,22], meaning that every single cell division, as well as each apoptotic event, has been described (Figure 4A).

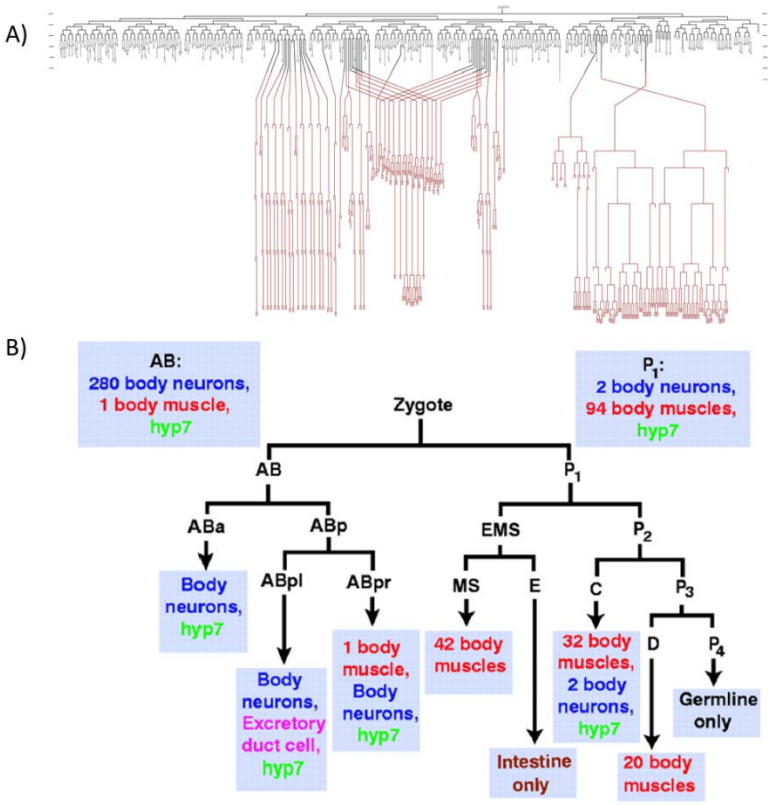


Figure 4. Cell lineages of C. elegans.

A) Embryonic (black) and post-embryonic (red) lineage of *C. elegans* hermaphrodites. Every single cell division and apoptotic event known is shown. Picture by Sulston and Horvitz, taken and adapted from www.wormatlas.org

B) Simplified schematic overview of the lineage of *C. elegans'* hermaphrodites, including lineage nomenclature and types of tissues that originate from different lineages. Picture by John Yochem and Robert K. Herman[23].

After the haploid oocyte is fertilized with a haploid sperm, originating either from the hermaphrodite itself or from a male, a single-cell diploid embryo is formed. Following the first cell division, the resulting 2-cell embryo consists of an AB and a P_1 cell, which further proliferate and give rise to several other subordinated lineages. The vast majority of body muscles, the germline and the intestine arise from the P1-cell, while almost all body neurons and the majority of the epidermis arise from the AB cell (Figure 4B). The adult epidermis (also: hypodermis) is mostly syncytial, which means that most of its cells are multinucleated - a result of several cell fusions during larval development. While most hypodermal cells consist of less than 10 nuclei in the adult hermaphrodite, the hyp7 cell consists of 139[24].

1.2.2 The *C. elegans* muscle system

C. elegans consists of two different types of muscles, striated (somatic) and nonstriated (single sarcomere) muscles[25]. The somatic (multiple sarcomere) muscles have attachment points to both the hypodermis and the cuticle. They consist of 95 so-called body wall muscles organized in four quadrants (Figure 5), 81 of which are born during embryonic development,

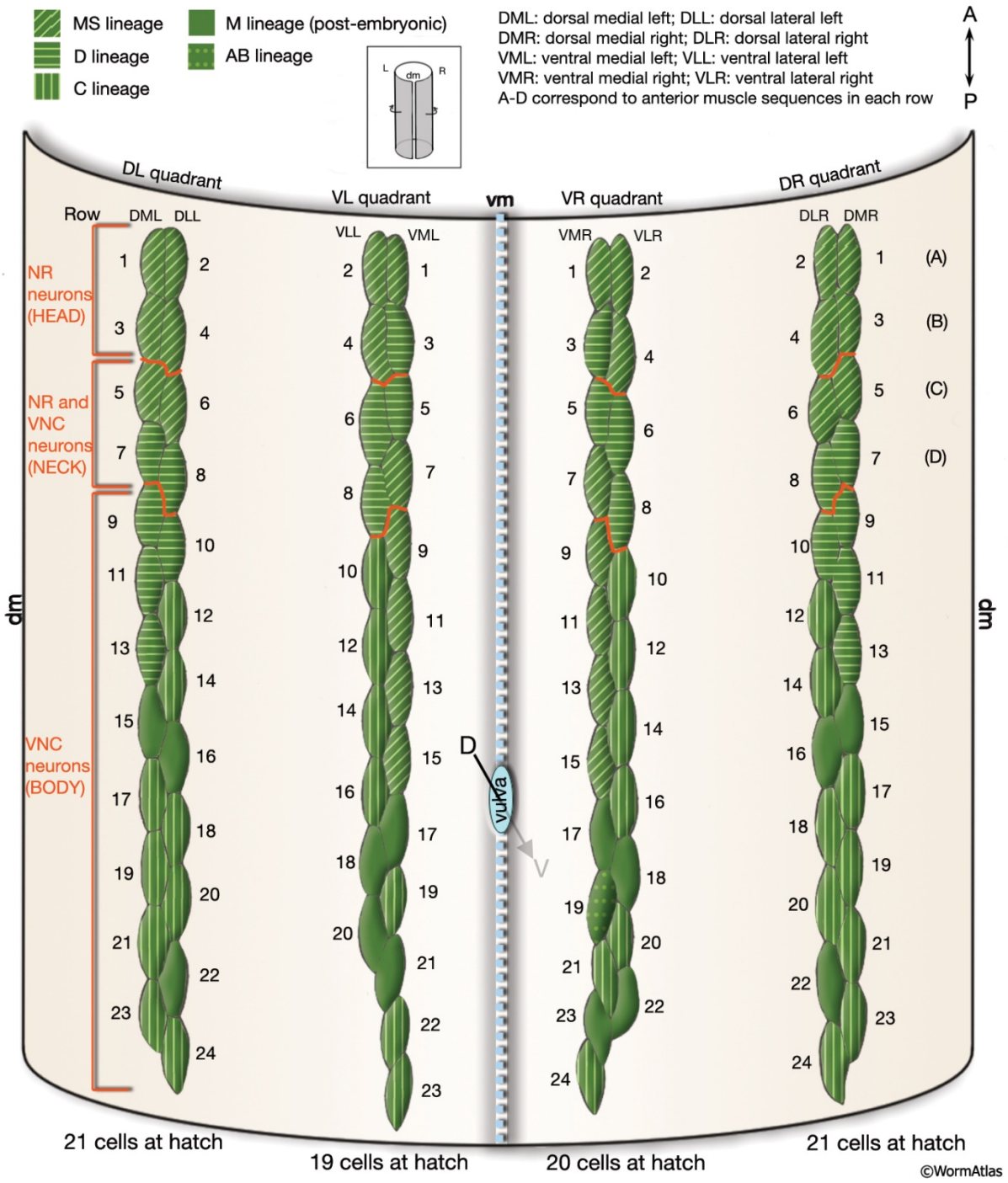


Figure 5. Arrangement of body wall muscles in *C. elegans*. The hermaphrodite is dissected along the dorsal midline and flattened, so that the ventral part in the middle is facing the viewer. Each quadrant contains 24 body wall muscle cells, except for the ventral left quadrant, which contains only 23. The lineage of which the different cells originated from is indicated. Post-embryonically born cells are shown in solid green. (dm) Dorsal midline; (vm) ventral midline. Picture taken from www.wormatlas.org

while 14 are generated post-embryonically during the transition from the L1 to the L2 larval stage. In contrast, the majority of nonstriated muscles have focal attachment points at their ends. They consist of 20 pharyngeal muscles, two stomatointestinal muscles, eight vulval muscles, eight uterine muscles, one anal sphincter muscle, one anal depressor muscle and the contractile gonadal sheath. Males obviously lack vulval and uterine muscles (as well as gonadal sheath), but have 41 specialized mating muscles instead. In general, all muscle cells are mononucleated, with the exception of the pharyngeal cells pm1-pm5, which fuse after hatching, resulting in a six nuclei pm1 syncytium and the five binucleated syncytial cells pm2-pm5.

Contractions and relaxations of the striated body wall muscle cells lead to the sinusoidal 'elegant' movement of the worm. *C. elegans'* somatic muscles send extensions to the dorsal and ventral nerve cord to connect to synapses, which is atypical, as for most animals muscles would receive axonal projections from motor neurons and not the other way around[26]. The first cloning and sequencing of a myosin gene happened in *C. elegans* in 1981 (*unc-54*)[27] and provided major insights into our understanding of myosin structure. There are many other so called *unc* genes, whose depletion leads to a phenotype of impaired movement (uncoordinated or Unc) and many of those are genes needed for proper muscle activity. On the other hand, not every myosin-encoding gene is termed an *unc* gene. One prominent example is *myo-3*, which encodes MHC A, the minor isoform of MHC (myosin heavy chain)[28].

myo-3 is primarily expressed in body wall muscles, but also in the somatic sheath, in enteric muscles, vulval muscles of the hermaphrodite and the diagonal muscles of the male tail[29]. Due to the equal and strong expression in all four quadrants of body wall muscles, the promoter of *myo-3* is widely used for reporters constructs (Figure 6).

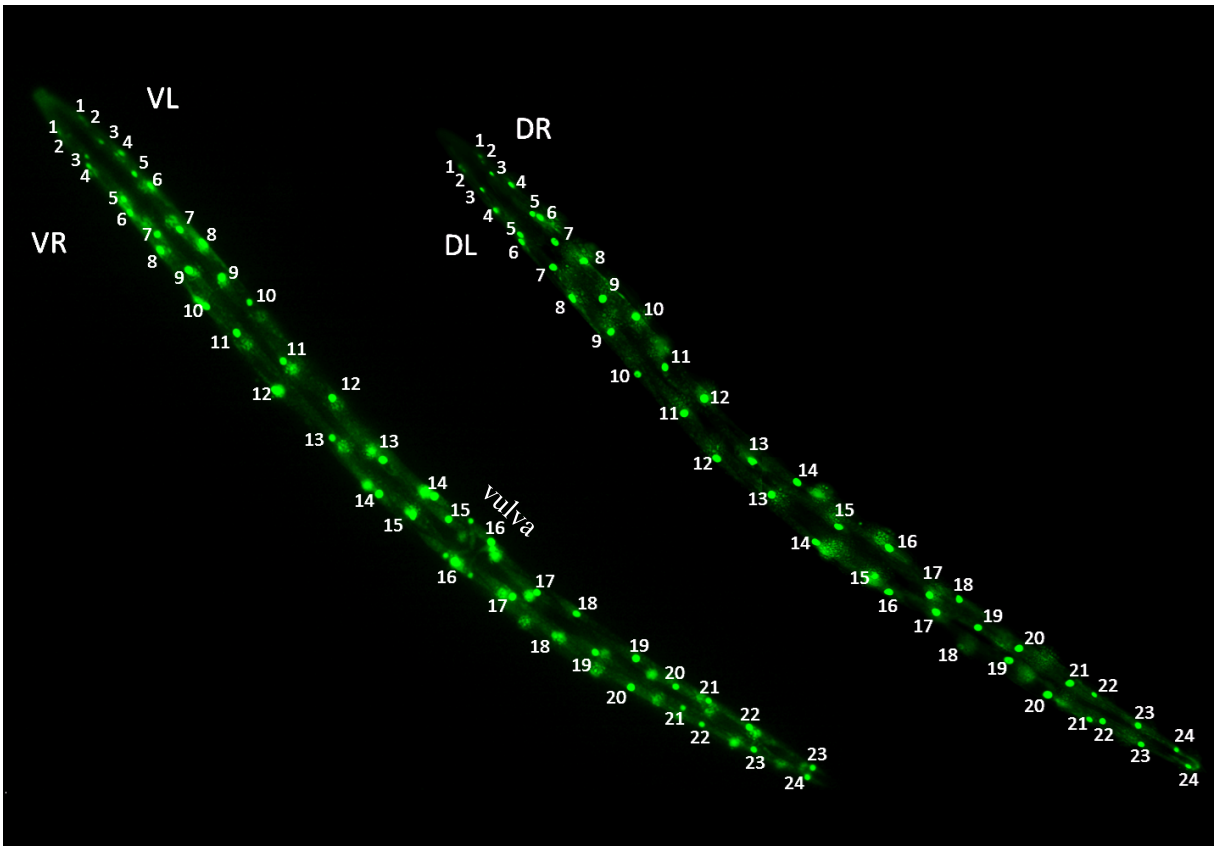


Figure 6. A *myo-3p::gfp::NLS* construct equally stains all four body wall muscle quadrants. Body wall muscle nuclei of each quadrant are consecutively numbered. Vulva nuclei on the ventral side are smaller in size and easily distinguishable from body wall muscle nuclei. (NLS) nuclear localization signal, **(**VR) ventral right, (VL) ventral left, (DL) dorsal left, (DR) dorsal right. Anterior is to the left.

1.2.3 The *C. elegans* gonad

In hermaphrodites, the gonad consists of two mirror-image U-shaped tubes, while in males it consists of a single U-shaped lobe only. The gonad consists of a somatic part and the germline, where oocytes (hermaphrodites only) and sperm (both hermaphrodites and males) develop[30]. The developmental roots of the germline can be found in early embryogenesis, where P blastomeres carry germline potential until P4 is born - the germline founder cell (Figure 7A). P4 is the first cell that gives rise to germ cells only and does not contribute to the soma at all.

Early germline specification is mediated by P granules, which are maternally provided ribonucleoprotein particles that are segregated to the P blastomeres. In the absence of P granules, germline development is severely impaired. Still, they are not sufficient for

specifying the germ cell fate. Besides P granules, the PIE-1 protein plays a crucial role in specifying the germline fate. PIE-1 is a maternally provided transcriptional repressor and keeps the early germline blastomeres in a transcriptionally silent state. MES-protein-mediated regulation of the chromatin state is believed to ensure the correct pattern of gene expression once the PIE-1 mediated transcriptional suppression ends[31].

Freshly hatched L1 larvae display four gonadal precursor cells: two primordial germ cells (Z2 and Z3) and two somatic gonad precursors (Z1 and Z4). Until mid-L1, those four cells are quiescent. Following this state, the somatic gonad develops in tandem with the germline (Figure 7B). The somatic precursors Z1 and Z4 proliferate to 12 cells by the end of the L1 stage: 10 proximal cells that will form the hermaphrodite somatic gonad primordium in late L2 stage and two distal tip cells (DTCs) that are crucial for the proliferation of the germ line. By the early L3 stage, the gonadal arms turn and extend rapidly, while the germ line proliferates in response to the DTC and the somatic sheath.

At the late L3 stage, first germ cells start to enter meiosis and a meiotic zone develops, which is growing in size during L4 development. By the late L4 stage, the germline is subdivided into a distal mitotic zone, a transition zone, a meiotic zone and proximal zone where spermatogenesis is happening. GLP-1 signaling, in a response of the DTC, inhibits the entry into meiosis in the mitotic zone. The genetic regulation of meiotic entry is complex and includes post-transcriptional gene regulators such as GLD-1, GLD-2, FBF-1/2 and NOS-3.

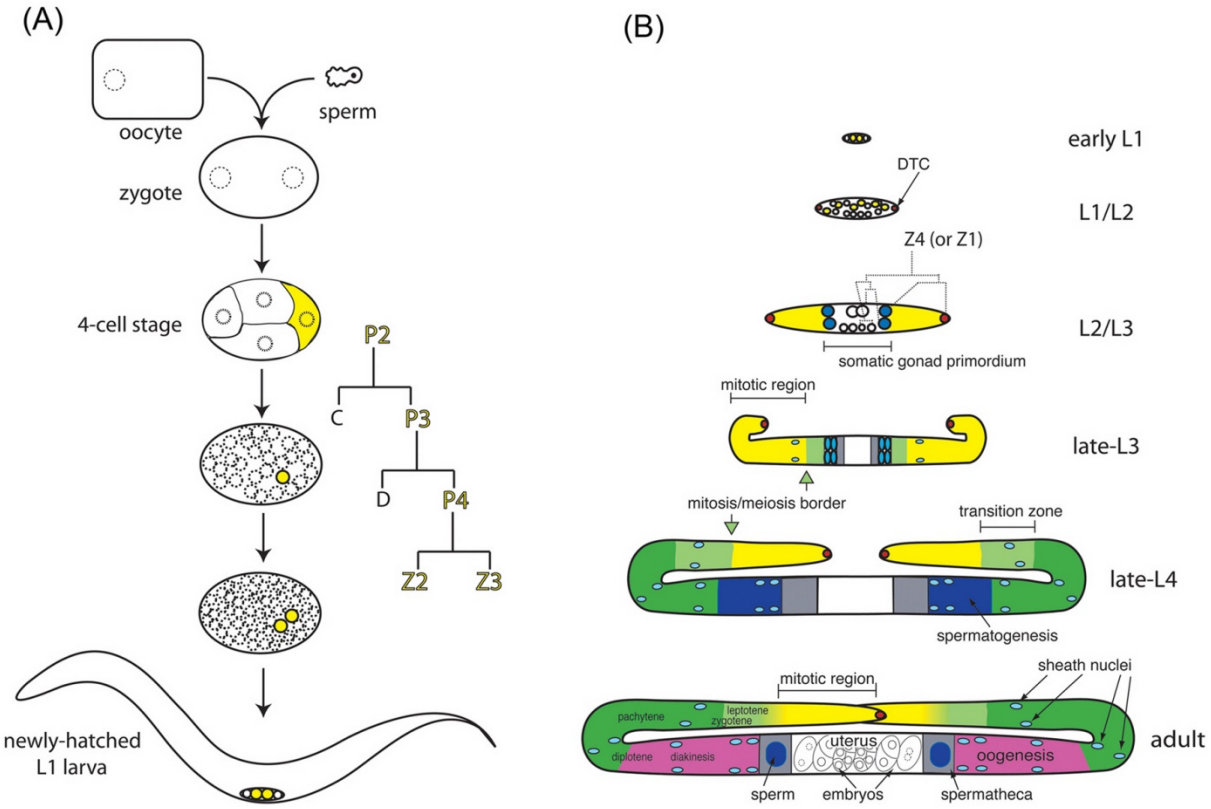


Figure 7. Gonadogenesis *in C. elegans* (A) Fertilization and the embryonic germline. Embryonic development initiates after fusion of oocyte and sperm. P cells carry germline potential and divide unequally until P4, which is the first cell that does not contribute to the soma and divides equally into Z2 and Z3. Germline lineages are depicted in yellow. (B) Post-embryonic hermaphrodite gonad development. Germline color scheme: yellow = mitotic region, light green = transition zone (early prophase of meiosis I), dark green = pachytene, dark blue = spermatogenesis, pink = oogenesis. In the late-L3 and late-L4 stages, the mitosis/meiosis border is indicated. In adults the mitosis/meiosis border is not sharp and mitotic and meiotic nuclei are interspersed at the border (indicated by a yellow/green color gradient). Somatic gonad color scheme: red = DTC, blue = sheath/spermatheca precursor cells, light blue = sheath nuclei, grey = spermatheca, white = uterus. Germline nuclei and their surrounding cytoplasm are often referred to as 'germ cells', although they are open to a core of shared cytoplasm (rachis) during most of their development. Picture by Jane Albert Hubbard and David Greenstein and taken from www.wormbook.org.

Spermatogenesis is finished at the adult stage, when sperm cells are stored in the spermatheca. In contrast to male gametes, that's the time point when oogenesis starts in hermaphroditic gonads and it continues throughout adulthood. Oocytes exhibit a huge cytoplasmic volume, as compared to germ cells, so that they can support early embryogenesis. Many female germ cells do not mature to oocytes, but undergo apoptosis. This is happening mostly at the gonadal turns (the so called death zone) and apoptotic germ cells are engulfed by gonadal sheath cells. Female germ cells that do not undergo apoptosis and mature to oocytes get either self-fertilized with sperm from the spermatheca or, alternatively, get fertilized by foreign sperm of a male. After fertilization, there is a short period of *in utero* development (2.5h at 20 °C), after which the developing embryos are laid as eggs.

1.3 Principles of genetic screens in *C. elegans*

C. elegans is highly amenable to genetic screens due to its short generation time, its hermaphroditic lifestyle and ease of use for both mutagenesis and RNA interference (RNAi) knock-down experiments. There is a wealth of gene-alteration techniques available, as well as ready to use RNAi libraries, overall making *C. elegans* a powerful tool for genetic screens.

1.3.1 Forward genetic screens

In a forward genetic screen a geneticist isolates mutants that display a phenotype of interest and tries to identify the mutated gene retrospectively. Often such mutations cause loss of function alleles, such as deletions, frameshifts, or premature STOP codons. Some of the first screens were performed using bacteria, as early as the 40s during the last century. For instance, Luria and Delbrück isolated bacterial mutants that were resistant to the bacteriophage T1, since they had lost the respective receptor[32]. The first mutagenesis screen in *C. elegans* was done in 1967 by Sydney Brenner, who published his results in 1974[12]. He isolated approximately 100 mutants that were viable and displayed visible phenotypes using chemical mutagenesis with Ethyl-Methane-Sulfonate (EMS). A schematic of an EMS mutagenesis can be found in Figure 8.

Besides EMS, several other chemical mutagens have been applied to the worm in order to perform forward genetic screens, including trimethylpsoralen with ultraviolet light (TMP/UV)[33], N-ethyl-N-nitrosourea (ENU)[34], formaldehyde[35], nitrosoguanidine (NTG)[36], diethyl sulfate (DES)[36], acetaldehyde[36], diepoxyoctane (DEO)[37] and diepoxybutane (DEB)[38]. EMS is not only the first mutagen ever applied to *C. elegans*, but still the most commonly used and potent one. It is an alkylating agent and most commonly adds an ethyl group to guanine to form O^6-ethylguanine[39]. In subsequent replications, this modified guanine inappropriately pairs with thymine[40], causing G/C to A/T transitions, which often leads to the generation of stop codons[41]. Such mutations are typically strong loss-of-function or null alleles. Besides point mutations, about 13% of EMS lesions are deletions or other chromosomal rearrangements[42].

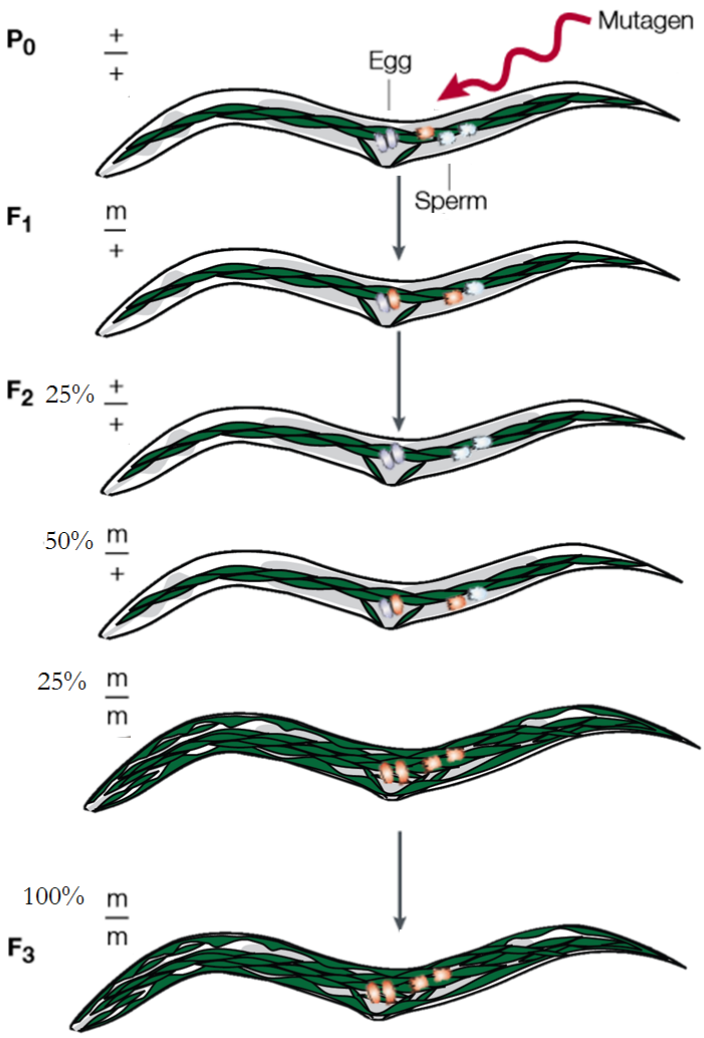


Figure 8. Schematic of a simple F2 mutagenesis screen. WT hermaphrodites are exposed to a mutagen and genes are randomly mutated in both somatic and germ cells (mutated germ cells are indicated in orange). For example, one sperm could be mutated for a specific gene, which depletion would result in ectopic GFP signals of a muscle-specific reporter. Fertilization of an egg by this sperm would result in a heterozygous F1 individual. Since this worm would be a self-fertilizing hermaphrodite, one-quarter of its F2 progeny would be homozygous for the mutation and display ectopic GFP. Such a F2 animal can be singled to a fresh plate and its F3 progeny can be inspected for homozygosity. Anterior is to the left. Image modified from Jorgensen & Mango[43] by Baris Tursun.

For the sake of completeness, it should be mentioned that apart from chemical mutagens, genome wide forward genetic screens in *C. elegans* can be also done by using high-energy radiation (ultraviolet light[44], ionizing radiation[45] and ^{32}P decay[46]) or transposons (*Mos1* transposase[47–49]).

1.3.2 Reverse genetic screens

In a reverse genetics screen a geneticist knocks out or knocks down genes of interest and analyses the resulting phenotypes. For small-scaled screens, knock-outs can be done using targeted mutagenesis, which aims to generate lesions in a specific gene of interest. Different methods of targeted mutagenesis exist and, as the name implies, just modify the targeted locus. In order to do so, the DNA double helix has to be cut, which can be achieved in two different ways: either by a transposon excision or by an enzyme, both of which will be

discussed in a later section. Off-target effects might still occur though, but can be dealt with by outcrossing mutant strains (such as it is usually done upon random mutagenesis in a forward genetic screen discussed in the previous section). Large scale screens usually rely on the use of RNAi.

1.3.2.1 Targeted mutagenesis

Historically, point mutations, small deletions, and insertions (e.g., of GFP) were introduced into a *C. elegans* gene of interest using *Mos1* excision-induced Transgene-Instructed gene Conversion (MosTIC)[50]. A similar strategy, Mos1-mediated Deletion (MosDEL), can be used to introduce large deletions (up to 25 kb) at specific sites[51,52]. Mos1-mediated Single Copy Insertion (MosSCI), a related method which also uses homologous repair after transposon excision, is not a mutagenic technique per se, but should be mentioned here for the sake of completeness. It can be used to introduce single copies of a transgene into a specific genomic site[52,53].

Enzymatic methods of targeted mutagenesis include Zinc finger nucleases (ZFNs)[54], transcription activator-like effector nucleases (TALENs)[55], and the CRISPR/Cas system (reviewed by Wiedenheft et al.[56]), summarized in Figure 9. ZFNs and TALENs are relatively complicated to design and expensive, whereas the CRISPR/Cas system is not.

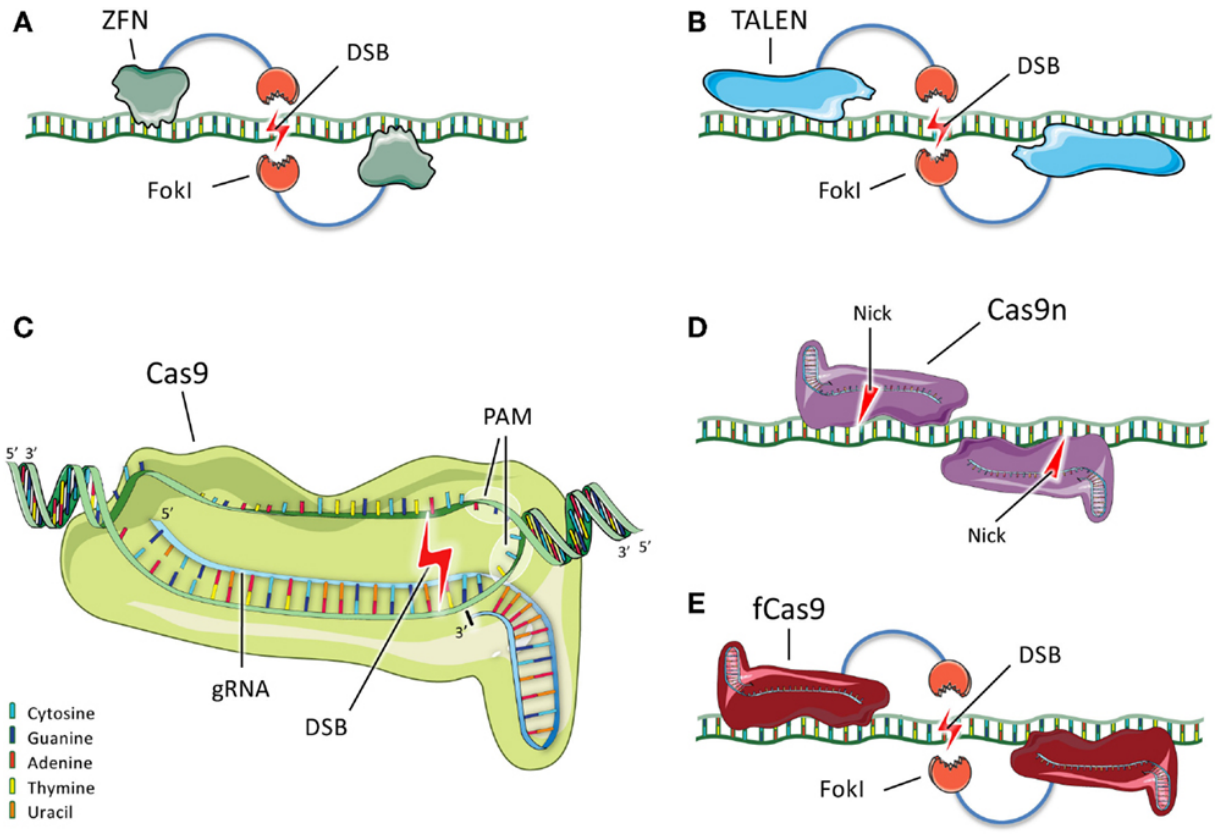


Figure 9. Enzymes to generate DSBs at specific loci: ZFNs, TALENs, CRISPR/Cas. (A) Two ZFNs are targeted to a locus of interest such that their FokI nuclease domains dimerize to generate a DSB. (B) Like ZFNs, TALENs are targeted to a specific locus, where they generate a DSB upon dimerization of their FokI domains. (C) In the CRISPR/Cas9 system, Cas9 is guided by a guide RNA that recognizes a 20-bp region in the genome. The Cas9/gRNA complex binds and introduces a DSB 3 bp upstream of the PAM motif (NGG). (D) Cas9 nickases (Cas9n) are mutant variants that possess a nickase activity and thus cannot cause a DSB on their own. Two nicks (one sense and one anti-sense) are needed to cause a DSB. (E) A similar strategy to avoid off-target effects uses catalytically inactive Cas9 fused with a FokI nuclease domain (fCas9). Only when two FokI nuclease domains dimerize, a DSB is caused. Image by Ott de Bruin et al.[57]

The discovery of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) system pushed genome editing to the next level. CRISPR/Cas is an adaptive immune system against viruses and plasmids (reviewed by Wiedenheft et al.[56]). The bacterial cell uses an RNA guide to target a Cas protein to a piece of foreign DNA in order to cleave it (Figure 9C). Different guide sequences to target foreign DNA can be stored in the bacterial genome (memory of the adaptive immune system). For cleavage *in vitro*, the minimal components required for the Type II CRISPR system are a synthetic single guide RNA and the Cas9 nuclease from *Streptococcus pyogenes*[58]. Target specificity is provided by a 20-bp sequence within the guide RNA. The only constrain is a Protospacer-Adjacent Motif (PAM) after the 3' end of the sequence, consisting of the trinucleotide NGG. Upon successful

targeting of Cas9 (mediated by a guide RNA), it introduces a DSB 3 bp upstream of the PAM site. Such a DSB may trigger non-homologous end joining (NHEJ), which is an error prone repair mechanism, potentially introducing indels. Alternatively, it might trigger homology directed repair (HDR), which is a repair mechanism relying on a (partially) homologous template and can therefore be used to insert a sequence of choice (e.g. a gene tag like GFP or FLAG). To reduce off-target effects, two genetically modified versions of Cas9, so called Cas9n, can be used together to target a specific locus (Figure 9D). Cas9n only possesses nickase activity, so one of the enzymes nicks the sense strand, while the other one nicks the anti-sense strand. Since both enzymes are needed simultaneously and they are each guided by their own guide RNA, this strategy greatly enhances target specificity[59]. Another strategy to avoid off-target effects uses two catalytically inactive Cas9 enzymes, that have a FokI nuclease domain attached (fCas9, Figure 9E). In order to cause a DSB, two FokI domains need to dimerize, which happens only if two fCas9s bind DNA in close proximity[60].

1.3.2.2 RNA interference

RNAi is probably the easiest and most straight forward method to do reverse genetic screens. Unlike targeted mutagenesis, it's very much feasible for large scale screens. It was discovered in *C. elegans*[61] and since then, a myriad of phenotypes could be analyzed based on gene knock-down experiments, where a transcript of choice gets depleted. By doing so, functional studies of genes were carried out in various organisms, including *C. elegans*, *Drosophila*, human cells and *Planaria*. RNAi is an endogenous gene silencing mechanism, where double-stranded RNA (dsRNA) targets messenger RNAs (mRNAs) of the same or similar sequence for degradation, cleavage or translational repression[61,62]. Whole genome RNAi libraries are available for several organisms, which enables silencing of most genes and thereby allows genome-wide loss-of-function-screens (Figure 10).

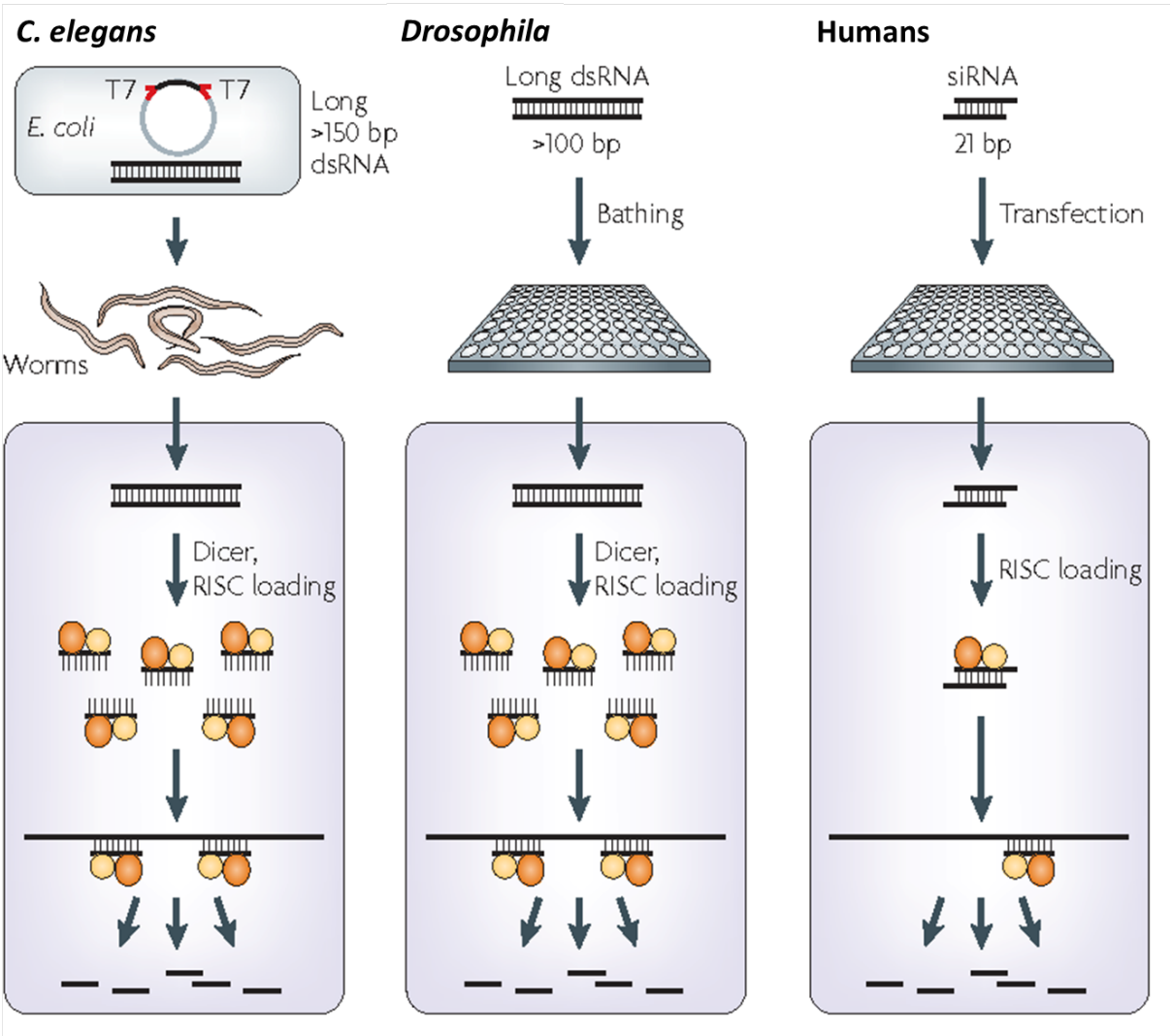


Figure 10. Overview of RNAi screening approaches used in different organisms. In *C. elegans*, RNAi is applied through dsRNA-expressing bacteria that are fed to the worms. In contrast, *Drosophila* cells are loaded with dsRNA by bathing. In both cases, the dsRNA is intracellularly processed and diced into small-interfering RNAs (siRNAs), which leads to a highly efficient knock-down, as several different siRNAs are generated from one molecule of dsRNA. In humans and other vertebrates, introduction of siRNA requires transfection. Usually several independent siRNAs targeting the same gene are used simultaneously to ensure a proper knock-down. Alternative methods include viral transfection of hairpin expressing constructs, which get processed by dicer *in vivo* and endoribonuclease-derived siRNAs (esiRNAs), which are generated from dsRNA *in vitro* using an enzyme such as RNAse III or dicer. RISC: RNA-induced silencing complex; T7: bacteriophage T7 promoter; Image by Boutros and Ahringer.[63]

RNAi libraries usually rely on the siRNA pathway, because siRNAs are specific to one target gene only. Besides the siRNA pathway, there is also a micro RNA (miRNA) pathway. Both siRNA and miRNA have similar physiochemical properties, but they have distinct functions and a different mechanism to trigger gene silencing (Table 1).

Table 1. Comparison of siRNA and miRNA

	siRNA	miRNA
Before being processed by Dicer	30 to over 100 nucleotides of dsRNA	70 to 100 nucleotides of precursor miRNA (pre-miRNA) with interspersed mismatches and hairpin structure
Structure	21-23 nucleotide RNA duplex with 2 nucleotides 3' overhang	19-25 nucleotide RNA duplex with 2 nucleotides 3' overhang
Complementarity	Fully complementary to target mRNA	Partially complementary to target mRNA(s), typically targeting the 3' UTR
mRNA target	one	several (could be over 100)
Silencing mechanism	Endonucleolytic cleavage of mRNA	Translational repression; Degradation of mRNA; Endonucleolytic cleavage of mRNA (rare, only for high complementarity)

Essentially, both of them are processed by Dicer and loaded into the RISC complex, but siRNAs have a single target only with a 100% sequence homology to a certain mRNA that gets cleaved, while miRNAs can have several targets with a <100% sequence homology and targeted mRNAs get either cleaved, degraded or their translation gets repressed.

It's important to mention here that in *C. elegans*, the primary RNAi response is enhanced by a secondary RNAi mechanism, which happens both up- and downstream of the initial dsRNA trigger[64–66]. The mechanism relies on RNA-dependent RNA polymerases (e.g. RRF-1 and EGO-1), that are recruited onto target mRNAs and synthesize secondary siRNAs that greatly increase the overall RNAi effectiveness. Secondary siRNAs are more abundant than primary siRNAs and associate with worm-specific Argonautes (WAGOs), which results in mRNA decay and co-transcriptional silencing (Figure 11).

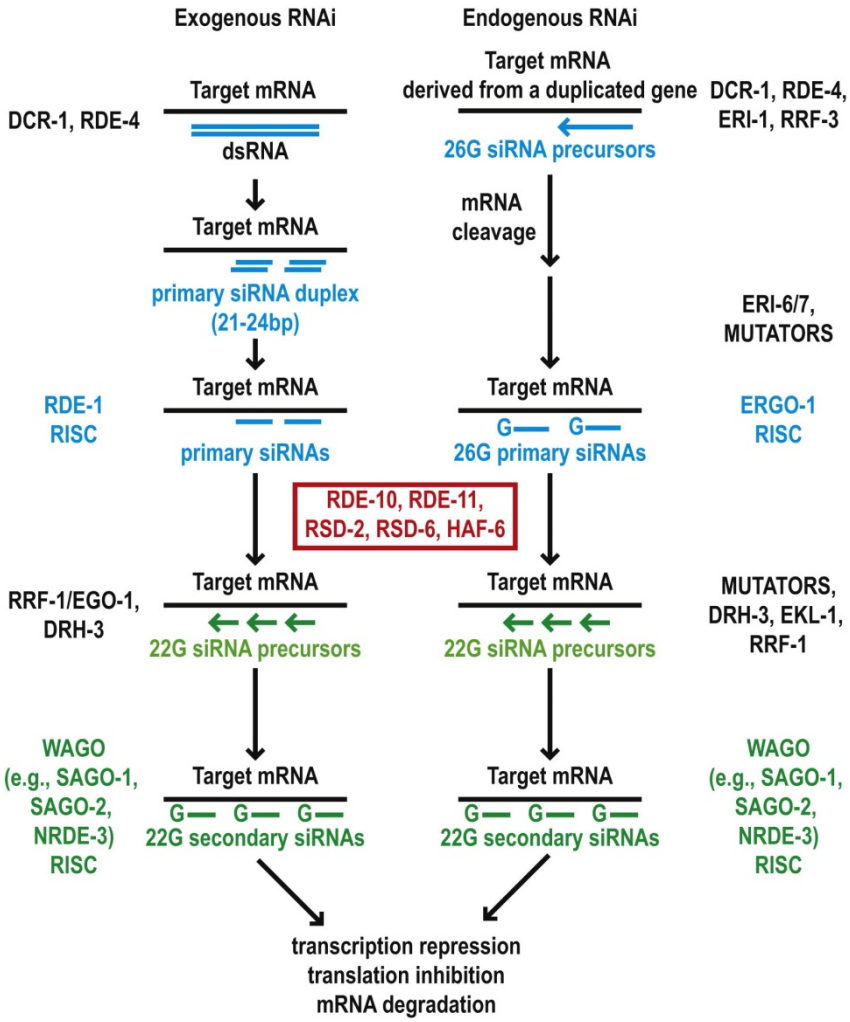


Figure 11 Primary and secondary RNAi responses for both exogenous and endogenous RNAi pathways. Red indicates the RDE-10/RDE-11 complex and other proteins that are required for the accumulation of a subset of WAGO class 22G siRNAs;
Blue indicates primary siRNAs, their precursors, and interacting Argonautes;
Green indicates secondary siRNAs, their precursors, and interacting Argonautes.

Image by Zhang et al.[67]

To apply RNAi in *C. elegans*, there are basically three different methods: injection, soaking and feeding[68]. The least labor intensive and cheapest method is feeding worms on bacteria that produce the desired dsRNA and, for that reason, it is also by far the most practiced method in *C. elegans* laboratories[69]. Results are a bit more variable than with the other two methods, but it is very reasonable to treat a large number of worms, both on solid or in liquid culture. Bacterial clones can be easily accessed from two major RNAi libraries. The Ahringer library was generated by cloning gene-specific genomic fragments between two inverted T7 promoters, thus containing both introns and exons[70,71]. It consists of 19762 clones, with an additional supplementary library of 3507 clones[72]. The Vidal library was generated by Gateway cloning of full-length open reading frame (ORF) cDNAs into a double T7 vector[73] and consists of 11511 clones, which target 10953 genes[74].

Generally, both forward and reverse genetic screens were, and still are, very powerful tools to get valuable insight as to how genes function and to gather information on what molecular events underlie a certain biological process.

1.4 Forward genetic screens targeted at isolating cell lineage mutants identified members of LINC complexes in *C. elegans*

Forward genetic screens that have been done in model organisms like *Drosophila* or *C. elegans,* contributed a lot to the initial insights into nucleo-cytoskeletal coupling LINC (Linker of Nucleoskeleton and Cytoskeleton) complexes. Genes that were identified in screens for a specific phenotype were often much later linked to other phenotypes, molecular pathways or protein networks. For instance, the SUN1 ortholog UNC-84 was already identified in 1980[75], but its identity as a SUN-domain protein was described almost 20 years later in 1999[76].
LINC complexes and especially the UNC-83/UNC-84 LINC complex are being discussed in the following sections.

1.4.1 LINC complexes

The nuclear lamina is connected to the cytoskeleton via different 'Linker of Nucleoskeleton and Cytoskeleton' (LINC) complexes with a variety of functions. LINC complexes are widely conserved over various phyla, which include organisms such as plants, slime molds, yeast, roundworms, fruit flies and mammals. LINC complexes cross the nuclear membrane and are composed of SUN and KASH domain-containing proteins, which interact in the perinuclear space between the inner and outer nuclear membrane. KASH proteins are located at the outer nuclear membrane and may interact with actin filaments, microtubules (via dynein and kinesin), intermediate filaments (via spectrin), centrosomes and other cytoplasmic organelles. SUN proteins are located at the inner nuclear membrane and are associated with both chromatin and nuclear lamins. Functions include nuclear movement and anchoring, moving meiotic chromosomes and telomeres and sensing mechanic stimuli[7–10].

1.4.2 The UNC-83/UNC-84 LINC complex

Both the KASH domain gene *unc-83* and the SUN domain gene *unc-84* have been identified in a forward genetics EMS screen in the early eighties[75]. The screen aimed to identify mutants with alterations of the normally invariant post-embryonic cell lineage. The initially observed phenotype for both *unc-83* and *unc-84* mutant animals was identical, namely a lower amount of nuclei in the ventral nerve cord as compared to WT animals. Interestingly, all seven mutant isolates of *unc-83* and *unc-84* were temperature sensitive, showing a more penetrant phenotype at elevated temperatures (25 °C instead of 20 °C or 15 °C)[77]. Sulston and Horvitz could link the lack of neuronal nuclei in the ventral nerve cord to a defect in P cell migration, which was temperature dependent. Mutant animals with a deficient ventral nerve cord displayed difficulties moving backwards (Unc phenotype). In addition, they found that *unc-83* and *unc-84* were required for the migration of certain hypodermal nuclei. During the embryonic development of the syncytial hypodermis, as well as during the postembryonic development of the ventral cord, certain cells develop protrusions superficial to body wall muscles. In WT animals, nuclei move along these protrusions. They observed that in *unc-83* or *unc-84* mutants, although these protrusions form normally, nuclei failed to move.

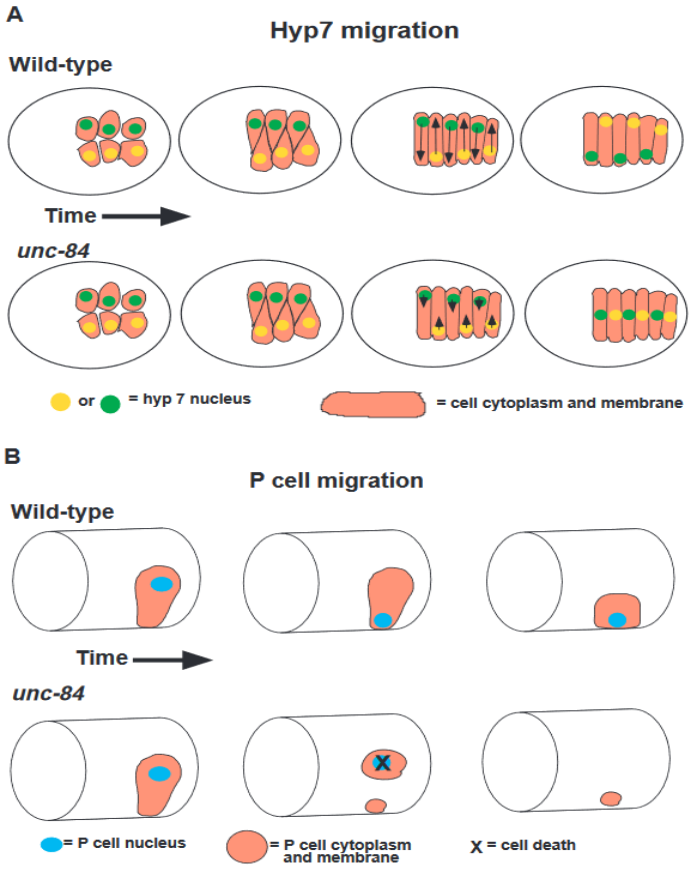


Figure 12. Nuclear migration defects in *unc-84* mutant animals. The graphic illustrates two different kinds of nuclear migration defects in *unc-84* mutant animals. (A) The process of six out of 17 hyp7 precursor cells undergoing cell elongation and nuclear migration are representatively shown. WT embryo is shown on top, *unc-84* mutant embryo at the bottom. The migration defect of the mutant is temperature independent. (B) P cell nuclear migration of one out of 12 P cells is shown. WT larva is shown on top, *unc-84* mutant larva at the bottom. The migration defect of the mutant is temperature dependent (normal migration at 15 °C, impaired migration at 25 °C). Anterior is to the left, ventral is downward. Image by Malone et al.[76].

Furthermore they describe how only the movements of two specific classes of hypodermal nuclei are affected by mutations in either *unc-83* or *unc-84*. After the initial isolation of those mutants, it took almost 20 years to shed some additional light on the subcellular localization and function of *unc-84*[76]. Malone and colleagues found out that UNC-84 localizes to the nuclear envelope and that it has a predicted transmembrane domain. Furthermore they describe migration defects of hyp7 precursor and P cells in more detail (Figure 12). While nuclear migration defects are present in both *unc-83* and *unc-84* mutants, they reported that *unc-84* mutants additionally show a nuclear anchoring defect in the hyp7 syncytium, suggesting that besides its role in nuclear migration, UNC-84 exhibits also a role in nuclear anchoring.

The lab of Daniel A. Starr confirmed these findings and identified the KASH domain of UNC-83, which enables binding to the SUN domain of UNC-84[78,79]. Together these two proteins form a LINC complex and cross the nuclear membrane, thus linking the cytoskeleton with the nuclear lamina. There are several other proteins involved that directly or indirectly bind to UNC-83, including NUD-2, LIS-1, DLC-1, BICD-1, EGAL-1, KLC-2 and UNC-116 (Figure 13). NUD-2/LIS-1 and DLC-1/BICD-1/EGAL-1 are two dynein-regulating complexes that get recruited by UNC-83. Besides dynein, UNC-83 also interacts with a second microtubule motor: kinesin-1. It consists of a heavy chain (UNC-116) and a light chain (KLC-2). Both microtubule motors have a distinct role in nuclear migration, as the Starr lab has shown in follow up analyses[80–82]. During embryonic hyp7 precursor nuclear migration, kinesin-1 is the major force that drives nuclei forward (towards the plus end of polarized microtubule), while dynein is responsible for bidirectional movements. Upon cytoplasmic roadblocks, dynein is able to move nuclei backwards for short stretches in order to bypass them (nuclear rolling). In contrast, during larval P-cell migration, dynein is the major force to move nuclei towards the minus end of microtubules. At the late L1 stage, P cells move to the ventral cord and the ventral surface is covered by the hyp7 syncytium. For this to happen, P-cell nuclei need to migrate from a lateral to a ventral position through a constricted space between body wall muscles and the cuticle. The constricted space is about 200 nm in width, which is about 5% of the pre-migration diameter of the nucleus. Failure of P-cell nuclear migration leads to Egl (egg laying deficient) and Unc (uncoordinated) animals due to the lack of certain vulval cells and motor neurons that originate from P-cells.

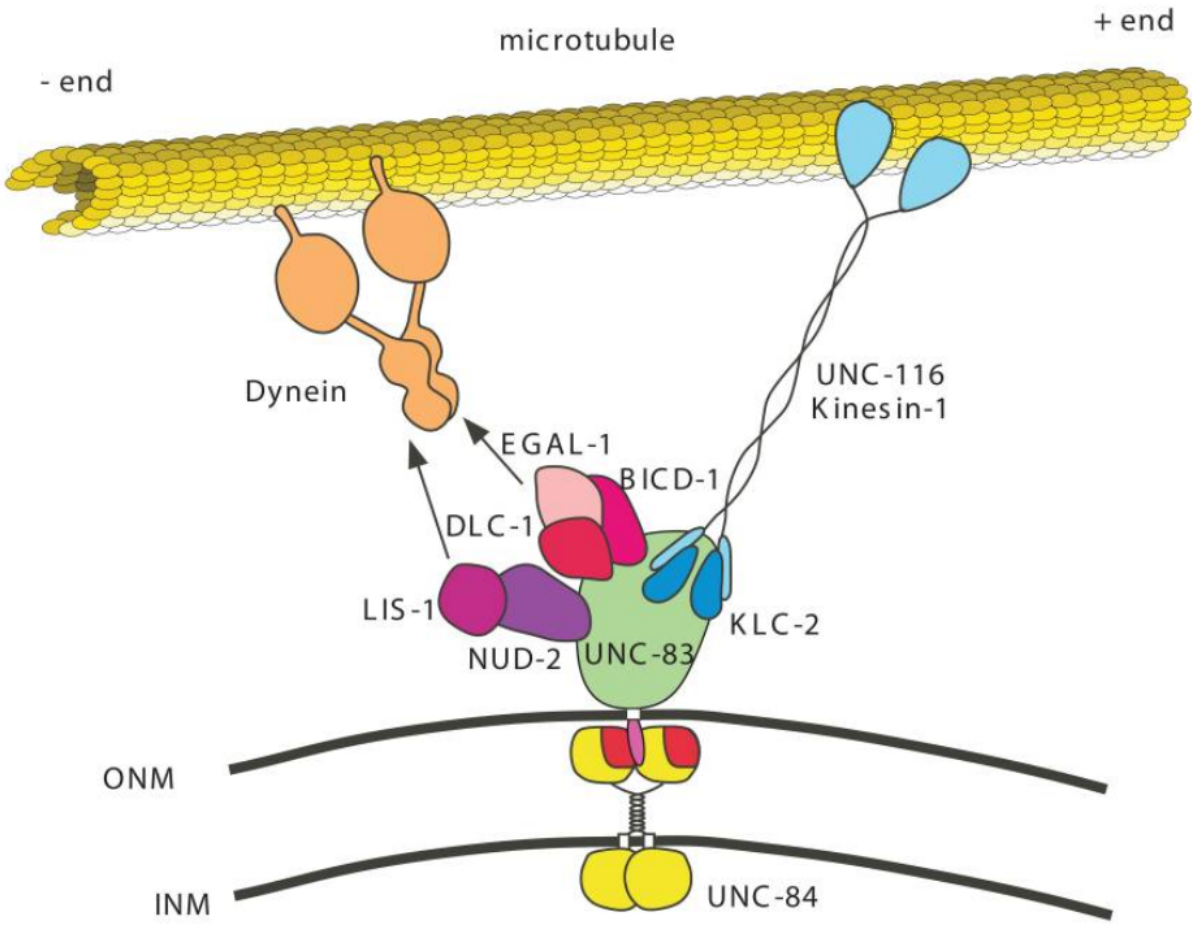


Figure 13. The UNC-83/UNC-84 LINC complex. UNC-84 is localized to the inner nuclear membrane (INM) and UNC-83 to the outer nuclear membrane (ONM).The KASH domain (pink) of UNC-83 interacts with the SUN domain (red) of UNC-84 in the perinuclear space. The cytoplasmic part of UNC-83 (green) interacts with the light chain of kinesin-1 (KLC-2, dark blue). The motor activity of the kinesin-1 heavy chain (UNC-116, light blue) moves nuclei toward the plus end of microtubules (yellow). UNC-83 also interacts with two dynein regulating complexes, NUD-2/LIS-1 (shades of purple) and DLC-1/EGAL-1/BICD-1 (shades of pink) which recruit motor dynein to the nuclear envelope to move the nucleus towards the minus-end of microtubules. Image by Fridolfsson et al.[79]

1.5 Reverse genetics identified reprogramming barriers in *C. elegans*

C. elegans is a powerful model organism to identify and study reprogramming barriers *in vivo*. The chromatin regulator LIN-53 (CAF-1p48/RBB7) for instance, was identified in *C. elegans* as a reprogramming barrier for the transdifferentiation of germ cells to neuron- and muscle-like cells in a reversed genetics screen[83,84]. Subsequently, the mouse ortholog of LIN-53, CAF-1, was shown to act as a reprogramming barrier in reprogramming somatic cells to

iPS cells[85], confirming the potential of using *C. elegans* as a system to identify evolutionary conserved reprogramming barriers.
Other reprogramming barriers that we have identified in our lab include HMG-3/4 and SPT-16, which form the FACT complex.

1.5.1 FACT (FAcilitates Chromatin Transcription) complex

FACT is a conserved heterodimeric complex composed of two subunits: SUPT16H (suppressor of Ty 16 homolog) and SSRP1 (structure-specific recognition protein 1) in mammals, Spt16 and Pob3 in yeast, as well as SPT-16 and HMG-3 or HMG-4 in *C. elegans*[1,86,87]. It is essential for maintaining stable gene expression by promoting transcription and functions by disrupting histone-DNA and histone-histone interactions, thus reorganizing nucleosomes. Besides its role in transcription, it was also shown to be involved in DNA repair[88] and replication[89–91]. It also promotes the integrity of chromatin structure on a genome-wide level, for instance, by suppressing cryptic transcription[87,92].

In a genome-wide reverse genetic RNAi screen in the nematode *C. elegans*, we identified FACT as a barrier for TF-mediated transdifferentiation[1] (Figure 14). We discovered that FACT in *C. elegans* exists in two different variants: one is germline-specific and consists of SPT-16 and HMG-3, while the other is predominantly somatic and consists of SPT-16 and HMG-4. FACT is required to maintain the germline and intestinal fate in *C. elegans.* Strikingly, its role as a reprogramming barrier is conserved, as its depletion enhances the reprogramming of human fibroblasts into iPSCs and neurons.

Identifying FACT as a reprogramming barrier was unexpected, because it is known as a positive regulator of gene expression, and previously described reprogramming barriers typically act as gene repressors[93]. Examples include the histone chaperone LIN-53 in *C. elegans*[83,84] (RBBP4/CAF-1P48 in mammals, section 1.4.2) and the Chromatin Assembly Factor 1 (CAF-1) in mice[85]. They promote gene silencing by introducing repressive chromatin marks, thereby blocking cell fate reprogramming. FACT is the first reported complex that promotes gene expression and still acts as a reprogramming barrier. Upon FACT depletion in human fibroblasts, the expression of many genes goes down, but some FACT-occupied genes show increased expression levels, suggesting that FACT might also have repressive functions.

Its combined role in promoting gene expression and preventing chromatin binding accessibility of certain TFs seems to be conserved in worms and humans and might be essential to maintain a specific cell fate program in both species.

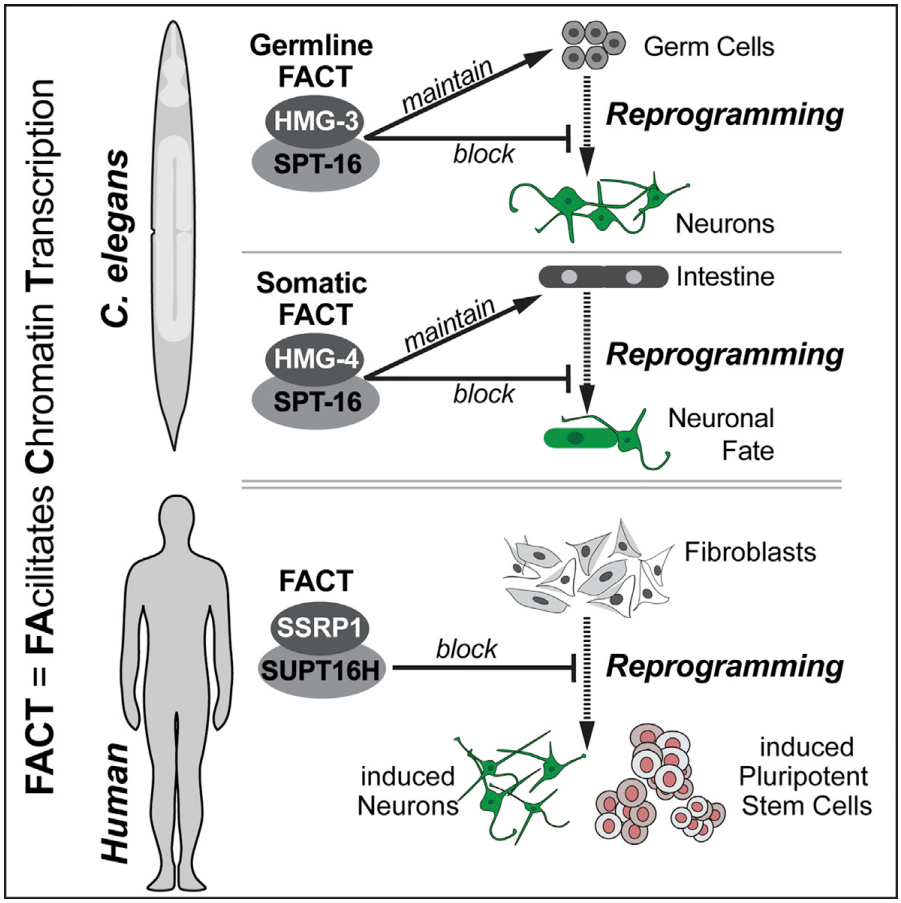


Figure 14. FACT as a reprogramming barrier in *C. elegans* and *H. sapiens*. The FACT complex consists of two different subunits. In *C. elegans*, there are two distinct variants, one which is germline specific, the other is somatic. It safeguards cell identities and acts as a reprogramming barrier in both *C. elegans* and human cell culture. Image by Kolundžić et al.[1].

1.6 Pseudogenes

Pseudogenes are DNA loci that share a sequence homology with a parental gene, but have lost some or all of their functionality[94,95]. There are several ways how pseudogenes can evolve, for instance, a non-essential gene might accumulate mutations that render it nonfunctional, but it can also happen through copy number variations, where DNA segments are duplicated or deleted. Although pseudogenes lost (some) functionality, they can still be functional, even if they aren't translated, and might perform regulatory functions similar to other kinds of noncoding RNA. In fact, many pseudogenes play a role in normal physiology and pathology such as cancer[96].

The two main characteristics that constitute a pseudogene are sequence homology to a parental gene and loss of some functionality. The former means that the parental gene and pseudogene diverged from a common ancestral sequence, while the latter characteristic means that any of the steps of proper gene expression, usually transcription, pre-mRNA processing, translation, and protein folding, is altered in a way that the original function of the gene is lost or changed. The most common reason for this would be a premature stop codon or a frameshift, which would prevent translation into a functional protein.

There are four different main types of pseudogenes: processed, non-processed, unitary and pseudo-pseudogenes.

• Processed (or retrotransposed) pseudogenes exist only in higher eukaryotes, especially mammals, as they originate from retrotransposition. Retrotransposition is the process of mRNA (or pre-mRNA) getting spontaneously reverse transcribed to DNA and being subsequently inserted into the genome[97,98]. They are typically characterized by a poly-A stretch at the 3' end (as a result of the poly-A tail of the original mRNA) and also by the lack of introns and a promoter.

• Non-processed (or duplicated) pseudogenes originate from a gene duplication event with the subsequent acquirement of mutations that would make the pseudogene lose its original function. Gene duplication might happen through homologous recombination, for example at repetitive sequences of misaligned chromosomes[99]. Non-processed pseudogenes usually share many characteristics with the parental gene like regulatory sequences or an intact exon-intron structure. Evolutionary gene duplications generate redundancy, and since the organism just needs one intact gene, mutations in the other would not affect its fitness. Since there is no natural selection on one of the copies, it gradually acquires mutations over time and transforms into a pseudogene.

• Unitary pseudogenes are the result of various mutations (such as indels or nonsense mutations) that would render a gene less- or non-functional, similar to non-processed pseudogenes. In contrast to non-processed pseudogenes, there was no gene duplication event prior to pseudogenization, and, as a consequence, the parental gene is lost. One example of a unitary pseudogene is the gene that encodes L-gulono-γ-lactone oxidase (GULO). In non-primate animals, GULO usually aids in the biosynthesis of ascorbic acid, but in humans and other primates, only a non-functional pseudogene variant of the gene, termed GULOP, exists[100].

• Pseudo-pseudogenes are genes that were originally classified as pseudogenes because of the

appearance of a premature stop codon in a mRNA sequence. Due to translational read-through, such a stop codon might be skipped and a small amount of functional protein might be produced[101,102].

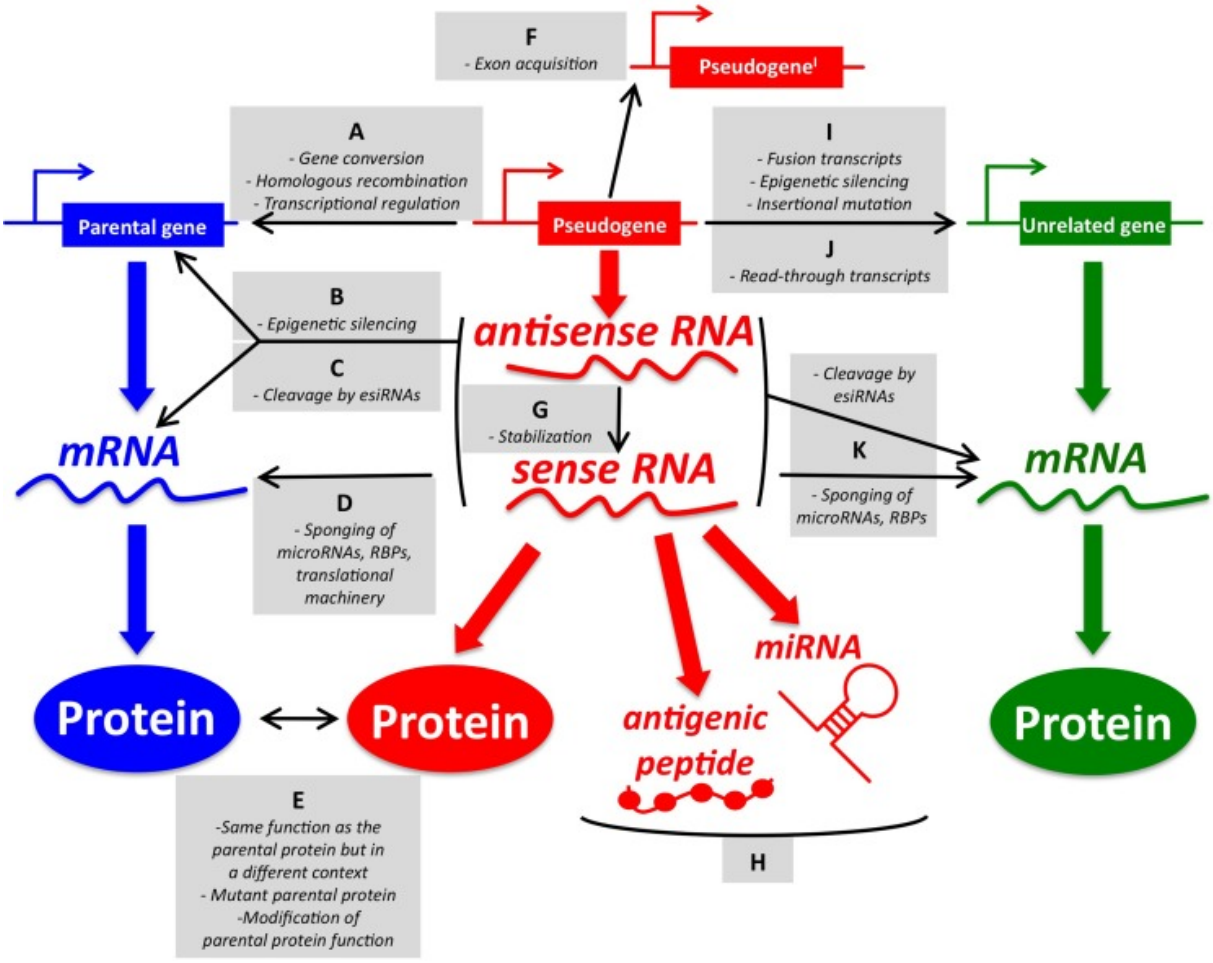


Figure 15. An overview of pseudogene functions. Pseudogenes might influence the parental gene (A-E), have functions unrelated to the mother gene (F-H) or may affect another gene (I-J). Image by Poliseno et al.[103]

Pseudogenes can be functional in many different ways[103] (Figure 15). For instance, the parental gene might get epigenetically silenced by sense or antisense transcripts of the pseudogene. Sense and antisense transcripts of the pseudogene might form dsRNA that gets cleaved into siRNAs and can downregulate the parental gene post-transcriptionally. Sense transcripts of the pseudogene might compete with the parental mRNA for the binding of microRNAs, RNA-binding proteins or the translational machinery ('sponging').

There are many other mechanisms that have been described of how a pseudogene can have a biological function. The interested reader might have a closer look at Figure 15 or might want to learn something about *F55A3.7*, a pseudogene that resembles an important part of this thesis and is introduced in the next section.

1.6.1 *F55A3.7* is a pseudogene related to *spt-16*

F55A3.7 is a previously uncharacterized pseudogene in very close proximity to its parental gene, the FACT subcomponent encoding gene *spt-16*. It is shorter than *spt-16*, oriented in the opposite direction less than 2 kb afar (Figure 16A) and shares a sequence homology of more than 90% with its parental gene (Figure 16B). Based on our data, *F55A3.7* gets spliced and consists of four exons, so it's likely that it is a non-processed pseudogene rather than a processed one. There is a complex substitution allele available, *F55A3.7(ok1829)*, which is a 1342 bp deletion with a 2 bp insertion spanning the second and third exon (Figure 16A).

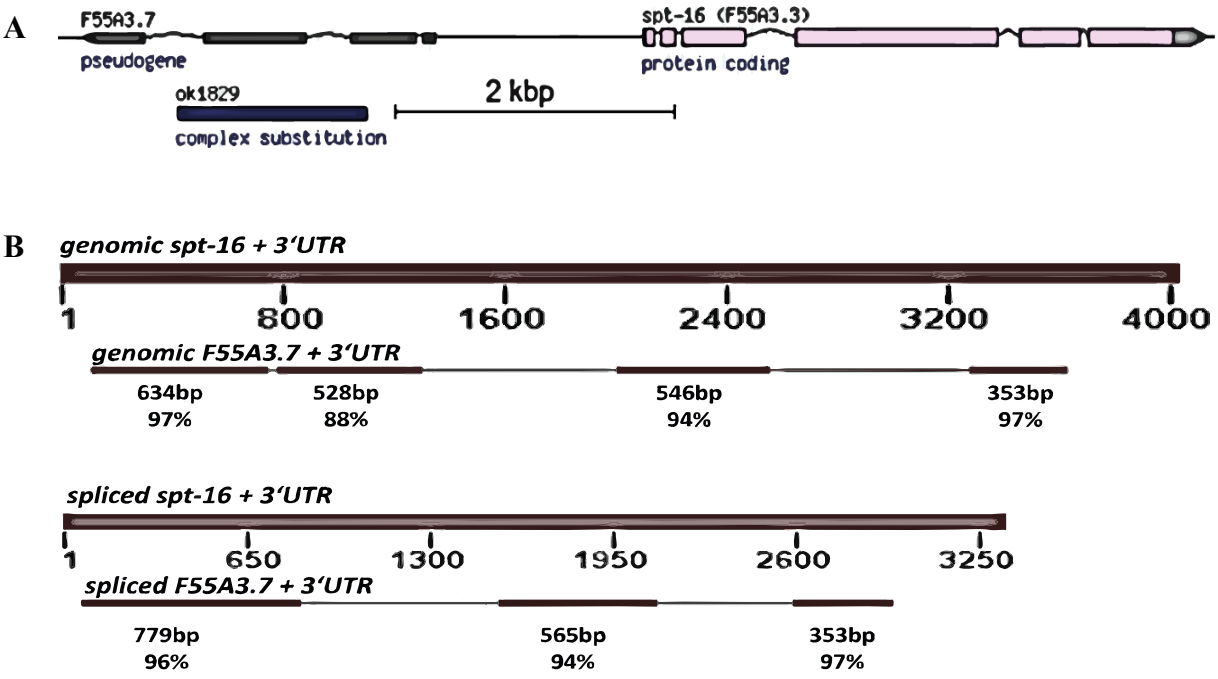


Figure 16. The pseudogene *F55A3.7* and its parental gene *spt-16*. (A) *F55A3.7* consists of four exons, is shorter than *spt-16* and oriented in the opposite direction less than 2 kb away. The *ok1829* allele is a variant, where a part of the sequence of the second exon and the whole sequence of the third exon are deleted. (B) Both the genomic and the spliced sequence of *spt-16* are highly similar to the genomic and spliced sequence of *F55A3.7*, based on a NCBI nucleotide blast.

We found that upon broad CHE-1 induction, *F55A3.7(ok1829)* mutant animals show a similar *gcy-5p::gfp* induction in germ cells as upon FACT depletion based on *hmg-3* RNAi (Figure 17 and Section 1.4.3). In F1 RNAi experiments, about 30% of *F55A3.7(ok1829)* animals displayed the *gcy-5p::gfp* induction phenotype (Figure 17A,B). Morphological changes like the shape of converted germ cell nuclei and the axo-dendridic like projections of *gcy-5p::gfp*

positive cells suggest a similar robust reprogramming, as we could show with the depletion of the FACT member *hmg-3*[1] (Figure 17C and Section 1.4.3).

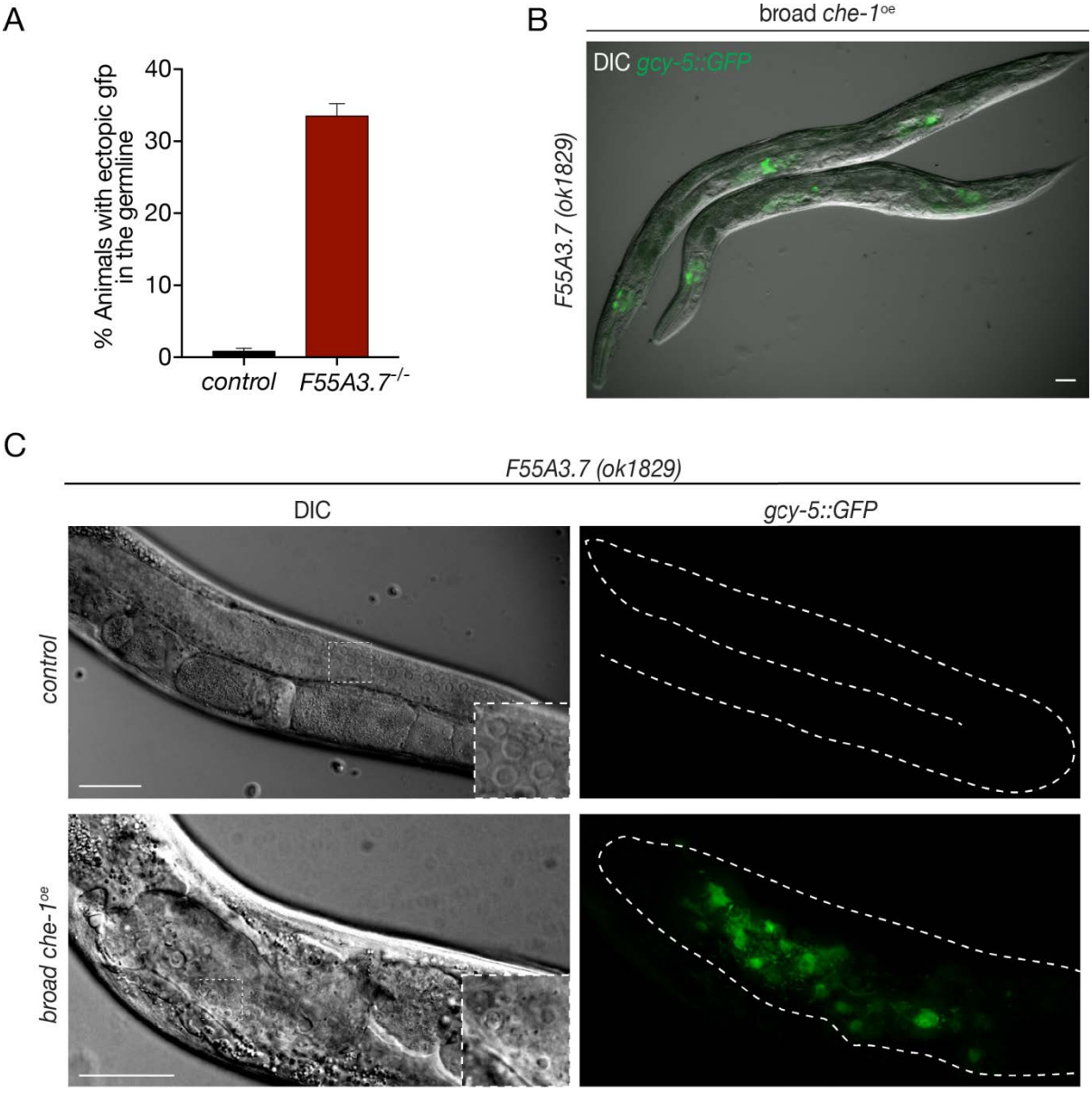


Figure 17. The pseudogene *F55A3.7* and its germ cells to neurons conversion phenotype upon depletion and broad CHE-1 overexpression. (A) F1 generation larvae (L4 state) of control animals (BAT28) and *F55A3.7(ok1829)* animals (BAT372) were quantified. Error bars represent SEM. (B) Representative image of GFP induction in the germline of *F55A3.7(ok1829)* mutant animals. Scale bar, 50 µm. (C) Converted germ cells underwent morphological changes. They lost their characteristic fried egg shaped nuclear morphology and adopted a speckled neuronal nuclear morphology (left) and, furthermore, displayed axo-dendritic-like projections (right). Dashed lines indicate the outline of gonads, dashed boxes indicate the magnification. Scale bars, 10 µm. Image by Ena Kolundžić[104].

1.7 Aim of this thesis

One aim of this thesis was to identify and characterize genetic factors that play a role in induced transdifferentiation by mis-expressing the TF HLH-1, which is the worm homolog of the myogenic bHLH TF MyoD. When mis-expressed, HLH-1 induces the muscle fate in early embryonic cells, but terminally differentiated cells in older animals are resistant to HLH-1-induced direct reprogramming. In order to identify mechanisms that antagonize HLH-1-induced reprogramming in ectopic tissues, I used a transgenic line allowing ectopic expression of the *hlh-1* gene in combination with a reporter for muscle fate (*myo-3p::gfp::NLS*). To do so, I developed a semi-automated high-throughput forward genetic screen combining EMS mutagenesis with the Biosorter system (Union Biometrica). Using this approach, I identified a mutant showing additional cells at the posterior end of the pharynx that expressed the *myo-3* reporter. I identified the mutated locus using whole genome sequencing and characterized the identified gene and the mutant phenotype further.

Additionally, I was also involved in characterizing the FACT complex, which was identified through a whole-genome RNAi screen conducted by my colleague Ena Kolundžić[1]. This reverse genetic screen, in contrast to the forward genetic screen that I performed, aimed at identifying factors that play a role in induced transdifferentiation by mis-expressing the TF CHE-1, a Zn-finger TF essential for terminal differentiation of glutamatergic ASE neurons. Interestingly, one of the FACT complex members, *spt-16*, is the parental gene of a previously uncharacterized pseudogene named *F55A3.7*. A putative null mutant of *F55A3.7,* combined with broad overexpression of CHE-1, showed a germ cells to neurons transdifferentiation phenotype similar to the phenotype observed after depleting the FACT member *hmg-3*. To our knowledge, this is the first example of a pseudogene whose depletion leads to the permissiveness of a certain tissue to be reprogrammed when challenged by a terminal selector TF. Due to this uniqueness, my aim was to characterize the pseudogene *F55A3.7* and to find a potential mechanism for how *F55A3.7* safeguards germline identity.

2 Results

2.1 A forward genetic screen reveals the requirement of the LINC member UNC-83 during positioning of body wall muscle nuclei in *C. elegans*

My initial aim was to identify and characterize genetic factors that play a role in induced transdifferentiation by mis-expressing the TF HLH-1, which is the worm homolog of the myogenic bHLH TF MyoD. When mis-expressed, HLH-1 induces muscle fate in early embryonic cells, but terminally differentiated cells in older animals are resistant to HLH-1-induced direct reprogramming. In order to identify mechanisms that antagonize HLH-1 induced reprogramming in ectopic tissues, I used a transgenic line allowing ectopic expression of the *hlh-1* gene in combination with a reporter for muscle fate, the transcriptional myosin reporter *myo-3p::gfp::NLS* (Figure 18A). To do so, I developed a semi-automated high-throughput forward genetic screen combining EMS mutagenesis with the Biosorter system (Union Biometrica, Figure 18B). Using this approach, I identified a mutant showing additional GFP-signals based on a *myo-3* reporter at the posterior end of the pharynx and named the mutant allele *bar18* (Figure 18C).

2.1.1 Additional muscle nuclei around the posterior pharyngeal bulb in *bar18* mutant animals are independent of ectopic *hlh-1* induction

With the initial semi-automated high-throughput EMS screen (Figure 18), I isolated BAT173, a strain containing a heat-shock inducible *hlh-1* transgene (*hsp-16.2::hlh-1*), a transcriptional myosin reporter for muscle fate (*myo-3p::gfp::NLS*) and the EMS induced *bar18* mutation. Since the phenotype of additional GFP positive nuclei around the posterior pharyngeal bulb was heat-shock independent and could be observed at different temperatures (15°, 20° and 25 °C), I outcrossed the *hlh-1* transgene using WT N2 worms to generate BAT197 (*ccls4251 [myo-3p::gfp::NLS], bar18)* which was used for further characterization. The independence of the phenotype from ectopic *hlh-1* induction already suggested that the mutated gene might not be a reprogramming barrier of induced transdifferentiation, but was causing a kind of developmental lineage defect.

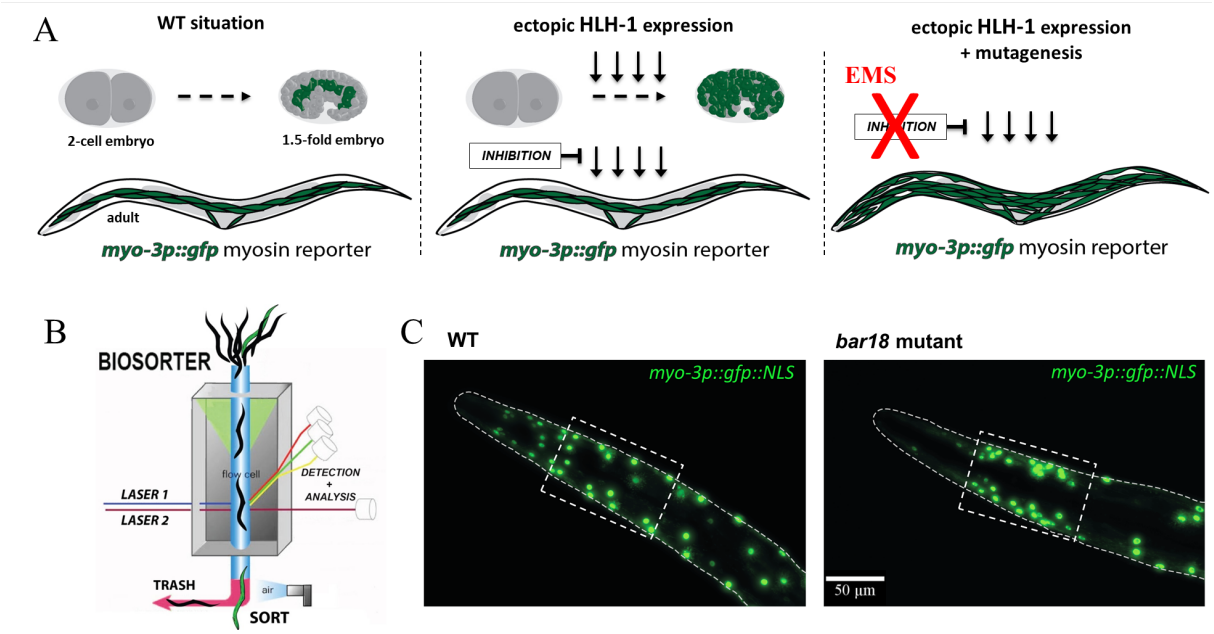


Figure 18. The mutant *bar18* was isolated by a semi-automated high-throughput EMS screen. (A) In a WT situation, *myo-3* active cells (green) can be found already in early embryonic stages and in adult worms (left). When an early embryo is challenged by ectopic HLH-1 expression, most cells will transdifferentiate into muscle like cells, unlike in adults, where cells are less plastic and inhibitory mechanisms safeguard the identity of already differentiated cells (middle). Upon removal of such inhibitory mechanisms by EMS mutagenesis, terminally differentiated cells in adult worms might also be reprogrammed to muscle like cells (right). (B) The BioSorter is a continuous flow system capable of analyzing and sorting 10-20 worms per second based on fluorescent intensity. (C) Compared to WT worms (left), *bar18* mutants (right) display an accumulation of body wall muscle nuclei around the posterior pharyngeal bulb (area indicated by dashed boxes). (NLS) nuclear localization signal; Anterior is to the left.

2.1.2 Whole-genome sequencing in conjunction with a SNP Mapping Strategy revealed that the mutated locus belongs to *unc-83*

To identify the locus that was affected by the *bar18* mutation, I used whole-genome sequencing in conjunction with a SNP Mapping Strategy[105] and a published CloudMap pipeline[106]. The principle is simple: I crossed the *bar18* mutant with a polymorphic *C. elegans* strain, the Hawaiian WT strain CB485, selected individual F2 progeny animals that showed the *bar18* phenotype and allowed them to self-fertilize. The F3 generations were pooled and submitted to whole-genome sequencing (WGS). The region of the phenotype-causing sequence variant had a lower density of polymorphic SNP markers and could therefore be identified in the WGS data. By doing so, I identified a premature STOP in the KASH-domain-containing gene *unc-83* (Figure 19A,B).

To test whether this mutation was causing the observed phenotype of accumulating body wall muscle nuclei around the posterior pharyngeal bulb, I performed rescue experiments, driving WT *unc-83* from an extrachromosomal array either under the control of the muscle-specific promoter *myo-3p* or the ubiquitous promoter *eft-3p* (Figure 19C, 20). Providing WT *unc-83* in the mutant background rescues the phenotype, thus confirming the identity of *bar18* as the premature STOP-causing mutation of *unc-83*.

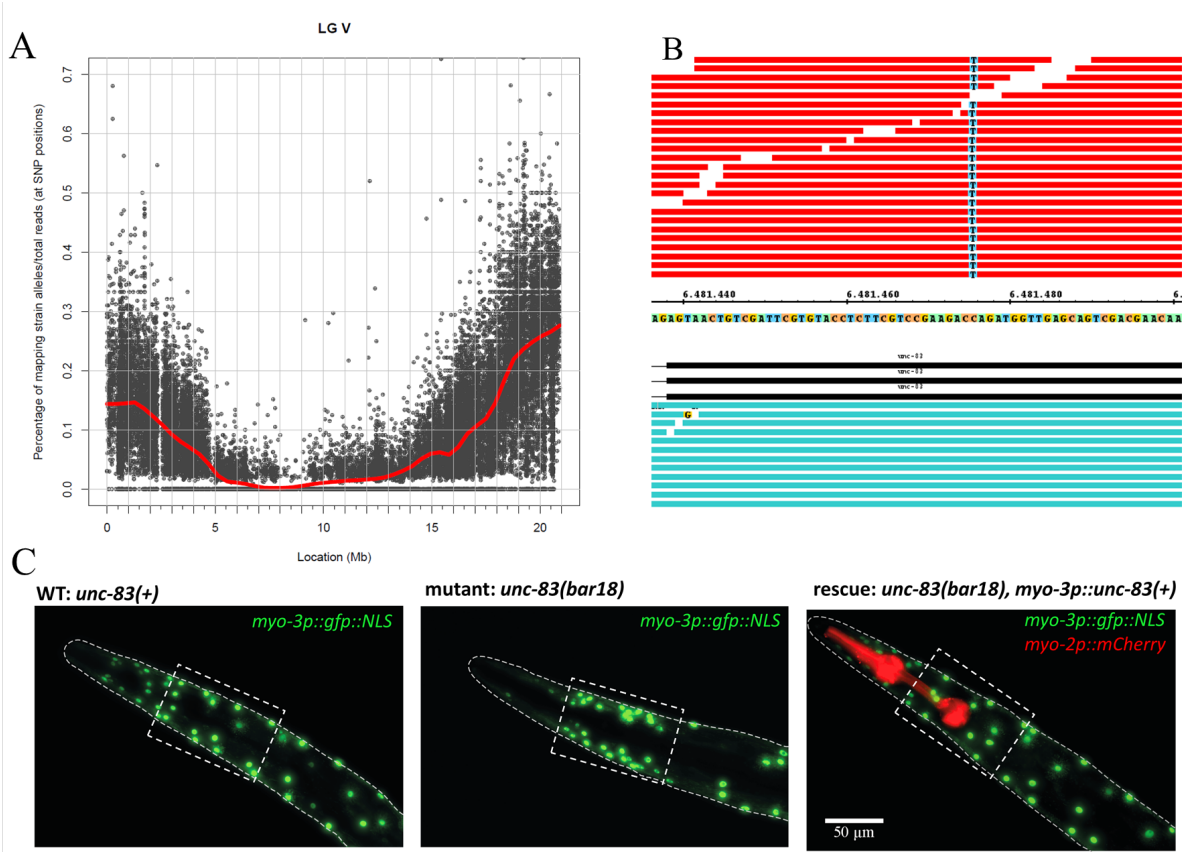


Figure 19. The *bar18* mutation causes a premature STOP in *unc-83* and has a cell autonomous effect on body wall muscle nuclei positioning. (A) Hawaiian SNP distribution reveals the position of *bar18* on chromosome V (see text for details). (B) Genome browser screenshot of WGS reads at a specific spot of the *unc-83* gene. A C>T conversion in *bar18* animals causes a premature STOP within the coding region of the gene. (C) Fluorescent images showing GFP-positive body wall muscle nuclei in WT (BAT661*), unc-83(bar18)* mutant (BAT1298, without the extrachromosomal rescue construct *barEx453*) and *unc-83(bar18)* mutant rescue animals (BAT1298, with the extrachromosomal rescue construct *barEx453*, driving WT *unc-83* from the muscle-specific promoter *myo-3p*). In the rescued animal, the pharynx expresses *myo-2p::mCherry* which is marking transgenic animals carrying *barEx453*. Dashed boxes highlight the area around the posterior pharyngeal bulb, where nuclei accumulate in the mutant. Anterior is to the left.

2.1.3 The accumulation of muscle nuclei around the posterior pharyngeal bulb in *unc-83(bar18)* mutant animals is a cell autonomous effect

Since driving WT *unc-83* from either the *eft-3* or *myo-3* promoter rescues the phenotype, I could not only confirm that the phenotype-causing mutation *bar18* indeed affects *unc-83*, but could also show that the effect is cell autonomous. Furthermore, I could phenocopy the accumulation of body wall muscle nuclei around the posterior pharyngeal bulb using a previously published *unc-83(ku18)* premature STOP allele (Figure 20). In addition, I tested the *unc-84(e1174)* deletion allele, which also shows this phenotype, supporting the assumption that a non-functioning UNC-83/UNC-84 LINC complex is responsible for the mis-positioning of muscle nuclei.

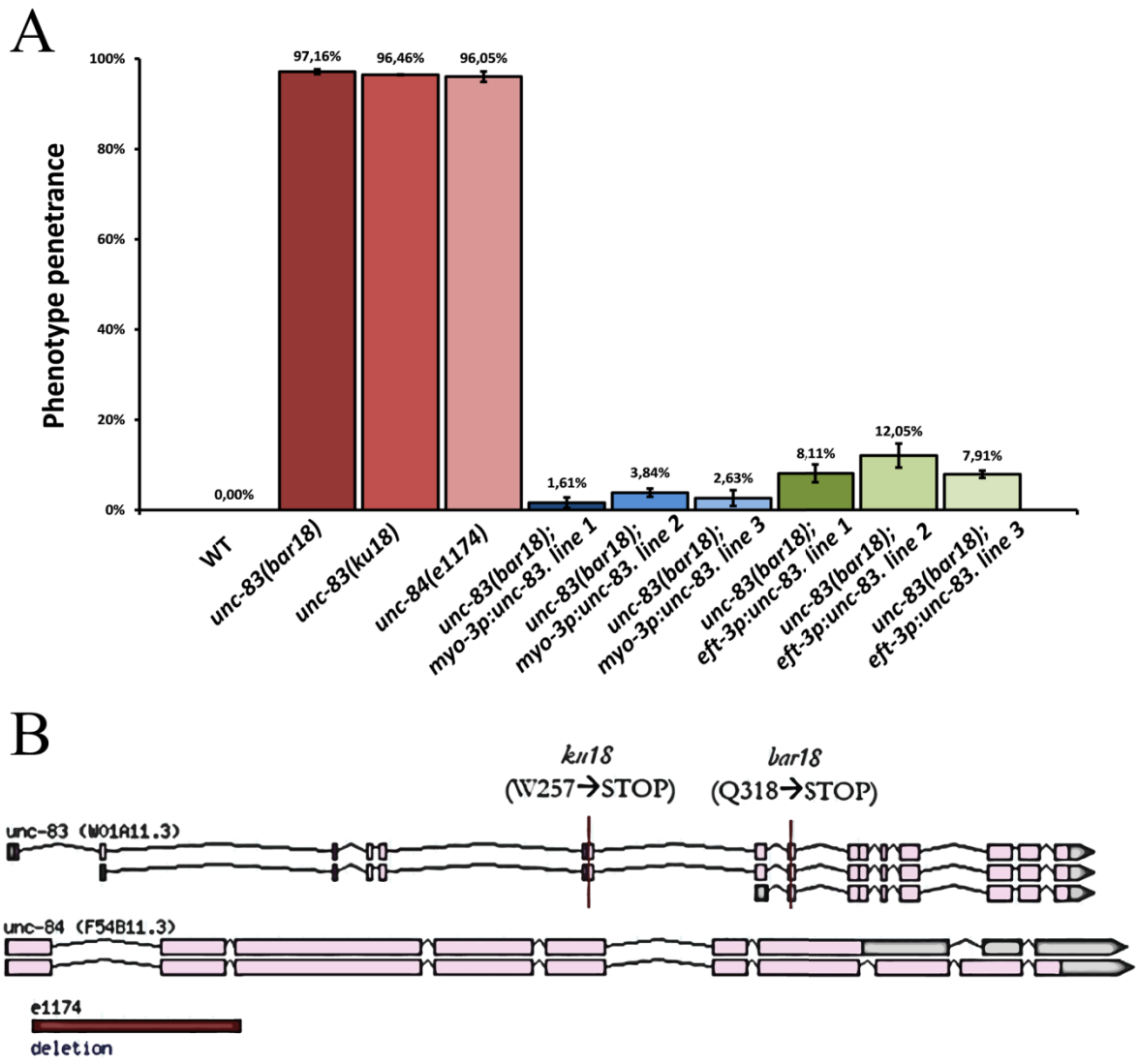


Figure 20. Different *unc-83/unc-84* alleles show the same phenotype of accumulating body wall muscle nuclei around the posterior pharyngeal bulb. (A) Phenotype penetrance of body wall muscle nuclei mis-positioning in WT (BAT661), *unc-83(bar18)* mutant (BAT197), *unc-83(ku18)* mutant (BAT1980), *unc-84(e1174)* mutant (BAT968), *unc-83(bar18)* muscle-specific rescued (BAT1298, BAT1906, BAT1907) and *unc-83(bar18)* ubiquitously rescued (BAT1300, BAT1908, BAT1909) animals. n ≥300 for each sample; 3 biological repeats; Error bars represent SEM. (B) Summary of *unc-83/84* alleles used in this study. Image adapted from genome browser of www.wormbase.org

2.1.4 Accumulated nuclei around the posterior pharyngeal bulb in *unc-83(bar18)* animals are not due to an overall increase in the amount of muscle cells

I quantified the amount of GFP-positive muscle nuclei in WT and *unc-83(bar18)* worms at different larval stages using the reporter construct *rrrSi261 [myo-3p::gfp::H2B]* (in contrast to the reporter described above, a single copy of *rrrSi261* is integrated into the genome thereby making it very dim, but more stable without any tendency for mosaicism). Surprisingly, the overall number of GFP-positive cells remained unchanged for all larval stages (Figure 21). At L1 stage, I counted 78.5 vs. 78.6 nuclei, at L3 stage I counted 94.7 vs. 94.6 nuclei and at L4 I counted 106.6 vs. 105.3 nuclei in *unc-83(bar18)* or WT animals, respectively. These results are comparable with previously reported numbers of muscle cells[13]: *C. elegans* has an invariant number of somatic cells, including 95 striated body wall muscles from the L2 stage onwards (81 in L1) and several other non-striated muscles, some of which are born at later larval stages.

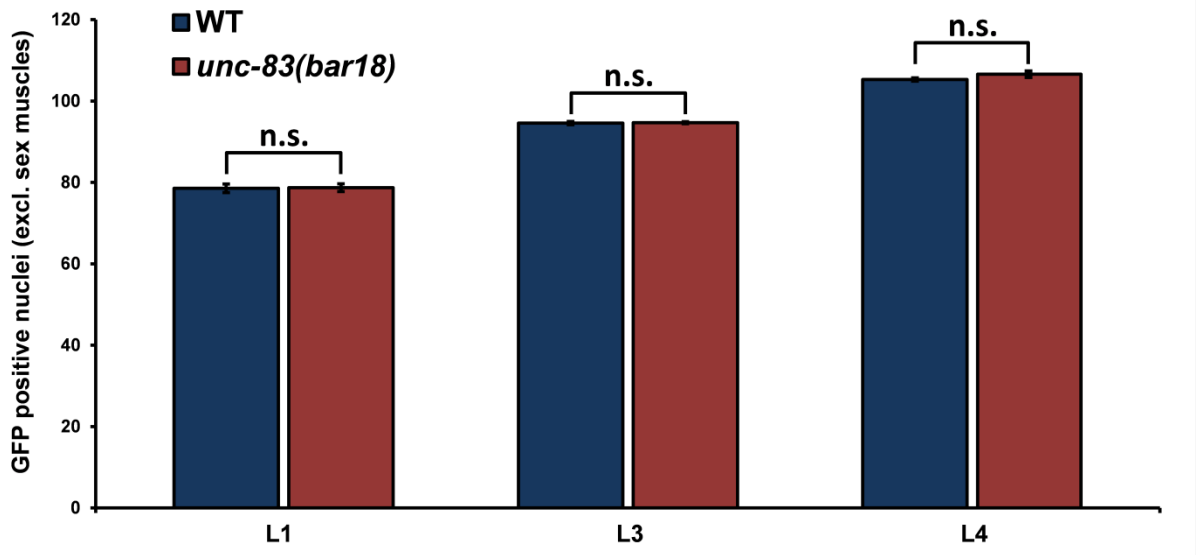


Figure 21. Amount of GFP-positive nuclei (excluding sex muscles) in WT (BAT1488) and *unc-83(bar18)* mutant (BAT1099) animals at the larval stages L1, L3 and L4. For L1 and L3 stages, these are exclusively body wall muscle cells. For the L4 stage, nuclei of other *myo-3p::gfp::NLS*-positive muscles, like somatic sheath or enteric muscles, contribute to the overall number of counted nuclei. n ≥10 for each condition; n.s. = not significant according to a student's t-test (p >0.05) between WT and mutant; Error bars represent SEM.

2.1.5 The UNC-83/UNC-84 LINC complex is required for proper body wall muscle nuclei positioning in *C. elegans*

Since the accumulation of muscle nuclei around the second pharyngeal bulb is not due to an increased number of muscle nuclei, it must be a positioning defect. In order to quantify the defect of body wall muscle nuclei in *unc-83(bar18)* animals, I sub-divided animals into 3 different regions (head, neck, posterior body) and quantified the amount of body wall muscle nuclei in WT and *unc-83(bar18)* animals (Figure 22). The neck region was defined as the region between the anterior pharyngeal bulb and the first pair of intestinal nuclei. The anterior region was defined as head and posterior was defined as posterior body. Overall, *unc-83(bar18)* animals displayed 36.1% less nuclei in the head region (7.8 vs. 12.3), 28.7% more nuclei at the neck region (28.3 vs. 22.0) and 3.1% less nuclei in the posterior body region (58.5 vs. 60.3) compared to WT animals. Our findings suggest that the nuclei accumulating in the neck region of *unc-83* mutants originate primarily from the head region.

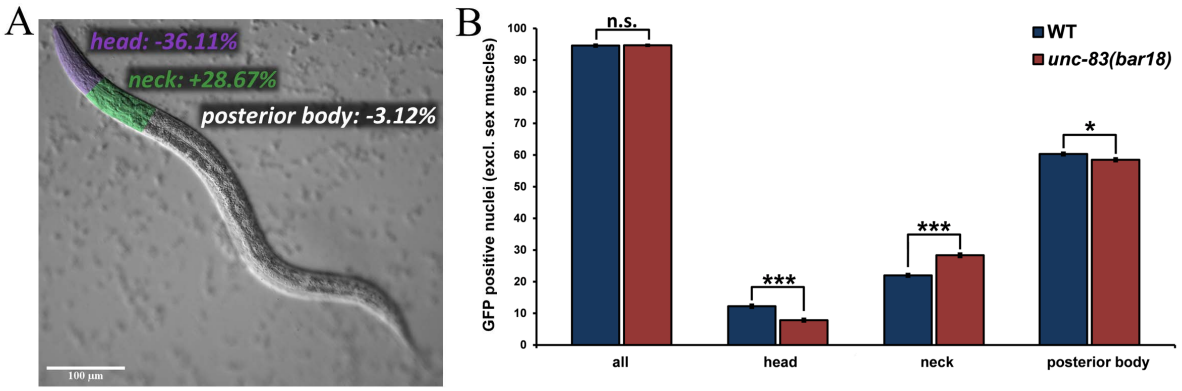


Figure 22. *unc-83(bar18)* animals display a positioning defect of body wall muscle nuclei. (A) Differential interference contrast (DIC) microscopic image of a L3 worm. The defined areas head (anterior part of the worm until after anterior pharyngeal bulb), neck (end of head region until after first pair of intestinal nuclei) and posterior body (end of neck region until posterior end of the worm) are highlighted. The differences in body wall muscle nuclei numbers between WT (BAT1488) and *unc-83(bar18)* mutant (BAT1099) worms are shown (based on the values shown in B).

(B) Distribution of body wall muscle nuclei in WT (BAT1488) and *unc-83(bar18)* mutant (BAT1099) L3 animals. There is no significant difference for the total number of body wall muscle nuclei. For the head and posterior body region, mutant animals display reduced amounts of body wall muscle nuclei, while the number of nuclei in the neck region is increased. Statistical significance based on student's t-test. n =12 for each condition; *p ≤0.02, ***p ≤0.0001; Error bars represent SEM.

In summary, I describe a so far uncharacterized phenotype of mis-positioned body wall muscle nuclei upon lack of a functioning UNC-83/UNC-84 LINC complex in *C. elegans*. Animals having mutations in *unc-83/84* display an accumulation of body wall muscle nuclei around the posterior pharyngeal bulb. The phenotype of *unc-83(bar18)* animals can be rescued by driving WT *unc-83* from either a muscle specific *myo-3* promoter or a ubiquitous *eft-3* promoter. My data suggests that this cell autonomous effect is primarily due to nuclei that are displaced from the head of the worm towards the neck region. Overall, these findings broaden our current understanding of this ubiquitously expressed LINC complex in *C. elegans*, which was so far described to ensure proper nuclei positioning in P cells, the intestine and *hyp7* hypodermal precursors, but not for muscle tissue.

2.1.6 Mutations of the LINC members *unc-83* and *unc-84* cause motility defects in *C. elegans*

As already outlined in the introduction, Sulston and Horvitz identified a temperature dependent Unc phenotype of *unc-83* or *unc-84* mutant animals that displayed difficulties moving backwards. This phenotype could be linked to a defect in P cell migration and the resulting reduction of neurons in the ventral nerve cord, which was more penetrant at elevated temperatures (25 °C instead of 20 °C or 15 °C)[77]. Since the positioning defect of muscle nuclei that I identified in *unc-83* and *unc-84* deficient animals is temperature independent, I was interested to see whether mutant animals would display any other motility defects, besides the difficulties of moving backwards at elevated temperatures, that I could link to this muscle specific phenotype.

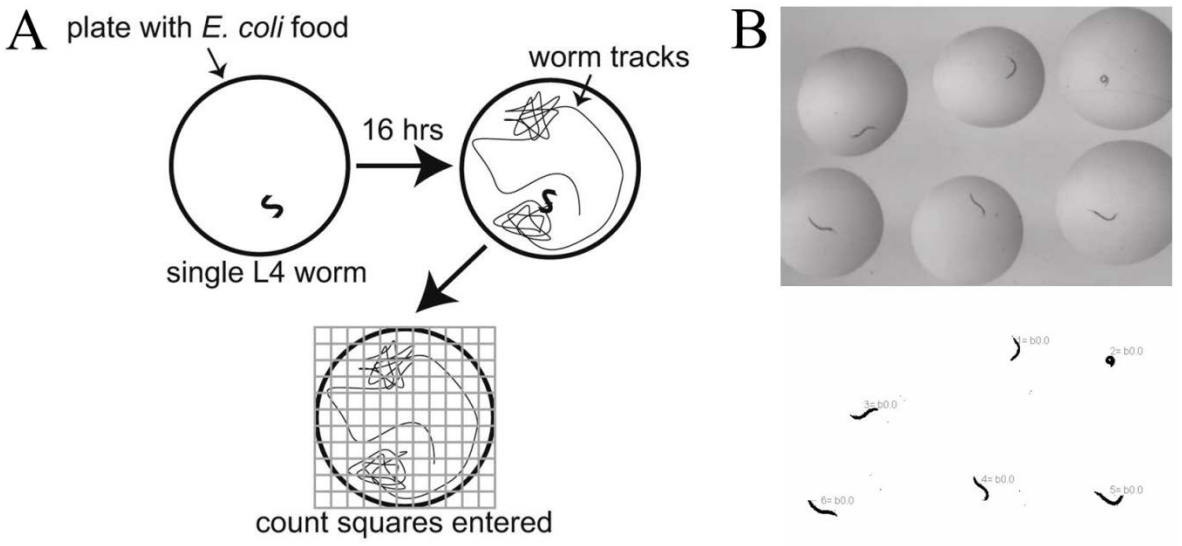


Figure 23. Motility assays. (A) An exploration assay measures exploration behavior based on movement across a bacterial lawn that has a grid applied to it[107]. Image taken from Flavell et al.[107]. (B) A swimming assay measures the amount of body bends an animal is doing while swimming in a droplet of liquid for 30 seconds. Videos are analyzed automatically using the ImageJ plugin wrMTrck[108].

I made use of two different assays to test this and did all experiments at 15 °C to uncouple any potential outcome from a lack of ventral nerve cord neurons due to impaired P cell migration. An exploration assay measures the motility based on movement across a bacterial lawn overnight (Figure 23A), while a swimming assay measures the body bends a worm is performing while being in a droplet of liquid (Figure 23B).

Both *unc-83(bar18)* and *unc-84(e1174)* animals performed worse than WT animals in an exploration assay (Figure 24). Compared to WT animals, who visited 2,83 squares/h on average, *unc-83(bar18)* animals visited 2,40 squares/h, while *unc-84(e1174)* animals visited only 2,26 squares/h. Interestingly, the impaired exploration behavior couldn't be rescued neither by driving WT *unc-83* from a muscle specific *myo-3* promoter (2,15 squares/h), nor from a ubiquitous *eft-3* promoter (2,32 squares/h, Figure 24).

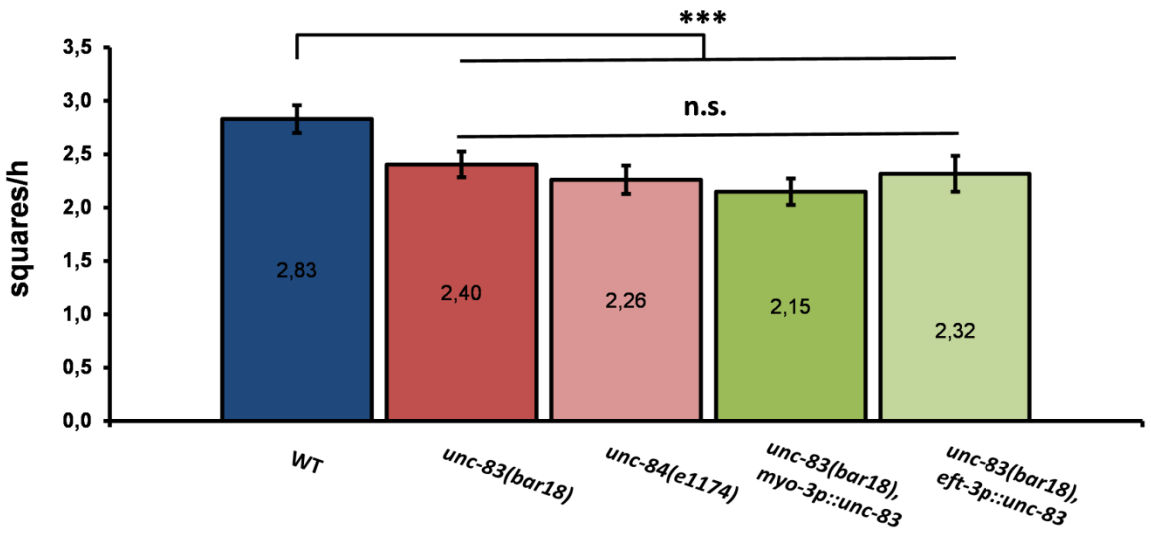


Figure 24. LINC deficient animals show decreased exploration behavior compared to WT animals. While WT animals explored 2,83 squares/h on average, both *unc-83(bar18)* and *unc-84(e1174)* explored significantly less squares. This phenotype couldn't be rescued neither by driving WT *unc-83* from a muscle specific *myo-3* promoter, nor from a ubiquitous *eft-3* promoter in an *unc-83(bar18)* background. WT animals visited significantly more squares than any of the other tested strains, while all other strains except WT showed no significant difference in their exploration behavior. n ≥39 for each condition. ***: p<0,01; n.s.: not significant.

To complement the exploration assays which measure long turn movements over a period of ~16 h, I performed swimming assays which measure short turn movements over a period of 30 sec (Figure 25). Like in the exploration assays, *unc-83(bar18)* animals showed an impaired movement compared to WT animals (1,57 body bends/sec vs. 2,03 in WT) and this defect could not be rescued by driving WT *unc-83* under the control of the muscle specific *myo-3* promoter (1,68 body bends/sec). In contrast to the exploration assays, driving WT *unc-83* from the ubiquitous *eft-3* promoter could fully rescue mutant animals in swimming assays (2,21 body bends/sec) that showed no significant difference to WT animal movements.

In summary, I could demonstrate so far uncharacterized motility defects of UNC-83 deficient animals at 15 °C, besides the already published impaired backward movements at elevated temperatures[77], using exploration and swimming assays. Two rescue lines, one driving WT *unc-83* from a muscle specific *myo-3* promoter and the other driving it from a ubiquitous *eft-3* promoter, could rescue the nuclei positioning defect in muscle cells (Figure 19C, 20A), but fail to rescue the exploration behavioral defects (Figure 24), suggesting that the exploration behavioral defect is not linked to mis-positioned nuclei in muscle cells. Motility defects identified by swimming assays were still present in the *myo-3p* rescue line, but were fully recued in the *eft-3p* rescue line (Figure 25), suggesting that the mechanisms of impaired

exploration and swimming are different and that the latter one might be connected to a tissue other than muscle (e.g. neurons).

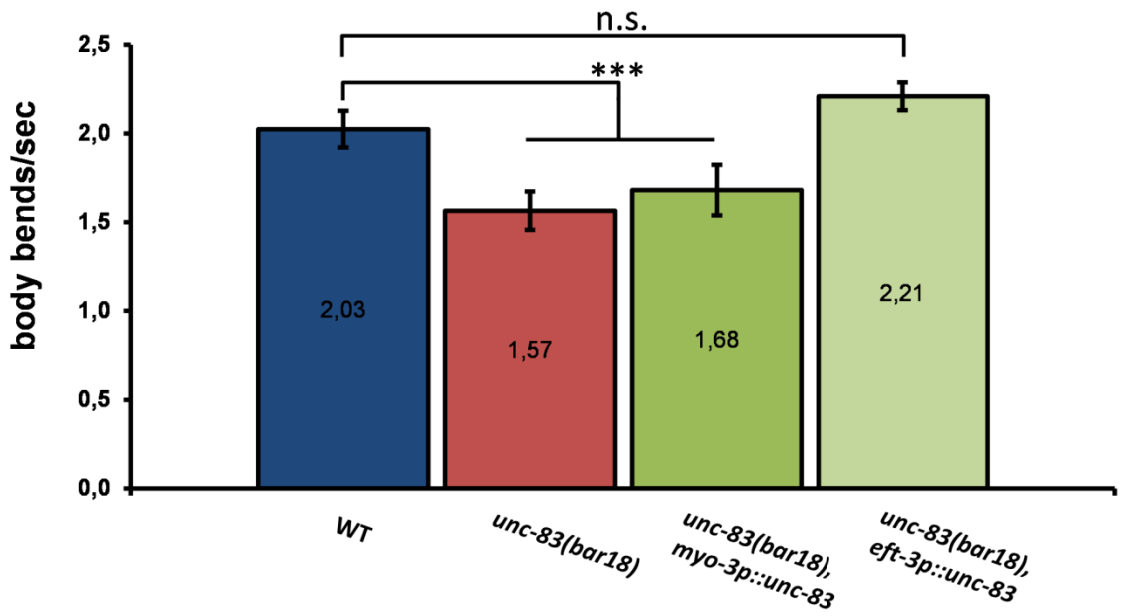


Figure 25. UNC-83 deficient animals show an impaired swimming behavior as compared to WT animals. WT animals show significantly more body bends/sec as *unc-83(bar18)* or *myo-3p* rescued animals. Driving WT *unc-83* from the ubiquitous *eft-3* promoter fully rescues the motility defect of *unc-83(bar18*) animals. n ≥25 for each condition. ***: p<0,01; n.s.: not significant.

2.1.7 Lack of the LINC member UNC-83 shortens the life span of *C. elegans*

I was interested to see whether the *unc-83(bar18)* mutation would cause other physiological phenotypes besides those that have been published previously (see section 1.3.2) and the positioning defect of muscle nuclei that I discovered[10]. One physiological defect that I could link to UNC-83 deficient animals was a shorter life span as compared to WT animals (Figure 26). Animals were kept at 15 °C until the late L4 stage to avoid defects in P cell migration[76,78,79] and were then put on plates containing 5-Fluoro-2′-deoxyuridine (FUDR) and shifted to 20 °C (see methods part for details). While WT animals displayed an average life span of 20,5 days, *unc-83(bar18)* mutants only lived for 16,6 days on average. Interestingly, driving WT *unc-83* from neither the muscle-specific promoter *myo-3p* nor from the ubiquitous promoter *eft-3p* could completely rescue the short life span phenotype, however, both rescue lines still showed a significant partial rescue (Figure 26).

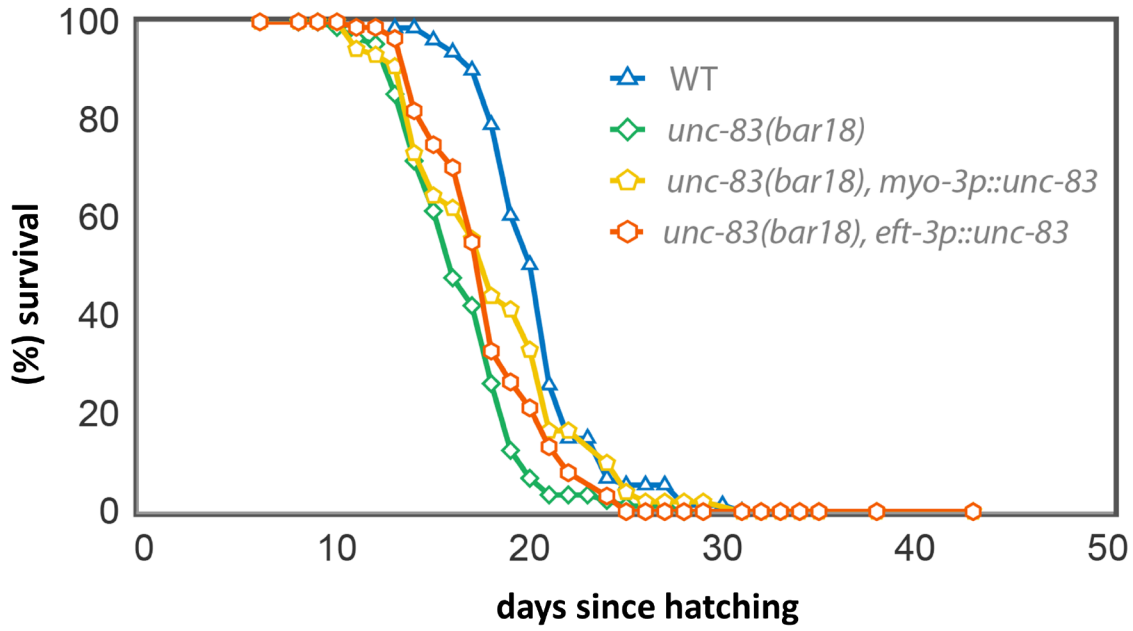


Figure 26. Mutants with a deficient UNC-83 show a decreased life span. While WT animals (blue line) had an average life span of 20,5 days, *unc-83(bar18)* mutants (green line) only lived 16,6 days on average. Driving an intact *unc-83* from a muscle specific *myo-3p* (yellow line) or a ubiquitous *eft-3p* (orange line), could only partially rescue the short life span (18,3 and 17,9 days respectively). 90 animals per condition, 3 biological replicates. P-values: WT vs. unc-83(bar18): $p \leq 0{,}01$; unc-83(bar18) vs. unc-83(bar18), myo-3p::unc-83: $p \leq 0{,}01$, unc-83(bar18) vs. unc-83(bar18), eft-3p::unc-83 = $p \leq 0{,}02$. Data analyzed using OASIS 2[109].

In order to see if the life span reducing effect of *unc-83(bar18*) was linked to the *daf-2* pathway, or to dietary restriction, which are both known to be involved in aging regulation[110,111], I tested the impact of two age-prolonging mutations on life span in a WT and *unc-83(bar18)* mutant background: *age-1(hx564)* (*daf-2* pathway, Figure 27) and *eat-2(ad465)* (dietary restriction, Figure 28). Both mutations increased the average life span of WT animals to a similar amount as they did in *unc-83(bar18)* animals (Figure 27, 28). WT animals lived 22,9 days on average when having the *age-1(hx564)* mutation (instead of 20,5 days, ≙ +11,7%), while *unc-83(bar18)* animals lived 18,7 days on average when having the *age-1(hx564)* mutation (instead of 16,6 days ≙ +12,7%; Figure 27). The *eat-2(ad465)* mutation increased the average lifespan of WT animals from 20,5 to 27,7 days (+35,1%), while it increased it from 16,6 to 24,2 days for *unc-83(bar18)* animals (+45,8%; Figure 28).

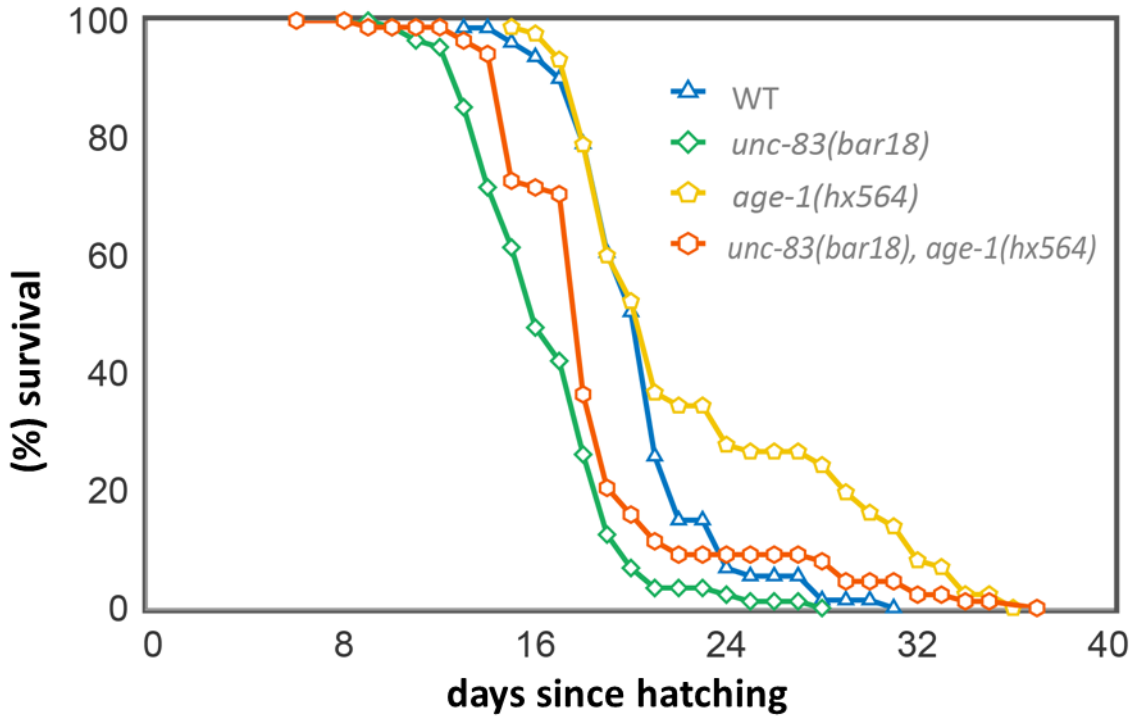


Figure 27. The life extending mutation *age-1(hx564)* increases the life span of UNC-83 deficient animals to a similar amount, as it does in WT animals. Having the *age-1(hx564)* mutation in a WT background (yellow line) significantly increases the average life by 11,7% as compared to WT (blue line), while having it in an *unc-83(bar18)* mutant background (orange line) significantly increases the average life span by 12,7% as compared to having the *unc-83(bar18)* mutation only (green line). 90 animals per condition, 3 biological replicates. P-values of all conditions: $p \leq 0{,}01$; Data analyzed using OASIS 2[109].

The life-shortening effect of *unc-83(bar18)* only partially suppresses the life-prolonging effects of *age-1(hx564)* and *eat-2(ad465)*, and both mutations extend the life span in a similar way as they do in WT worms, which suggests that they act independently of each other (they are non-epistatic to one another). My results support the hypothesis that the effect of lacking UNC-83 on *C. elegans'* life-span is unrelated to both the *daf-2* pathway and to dietary restriction.

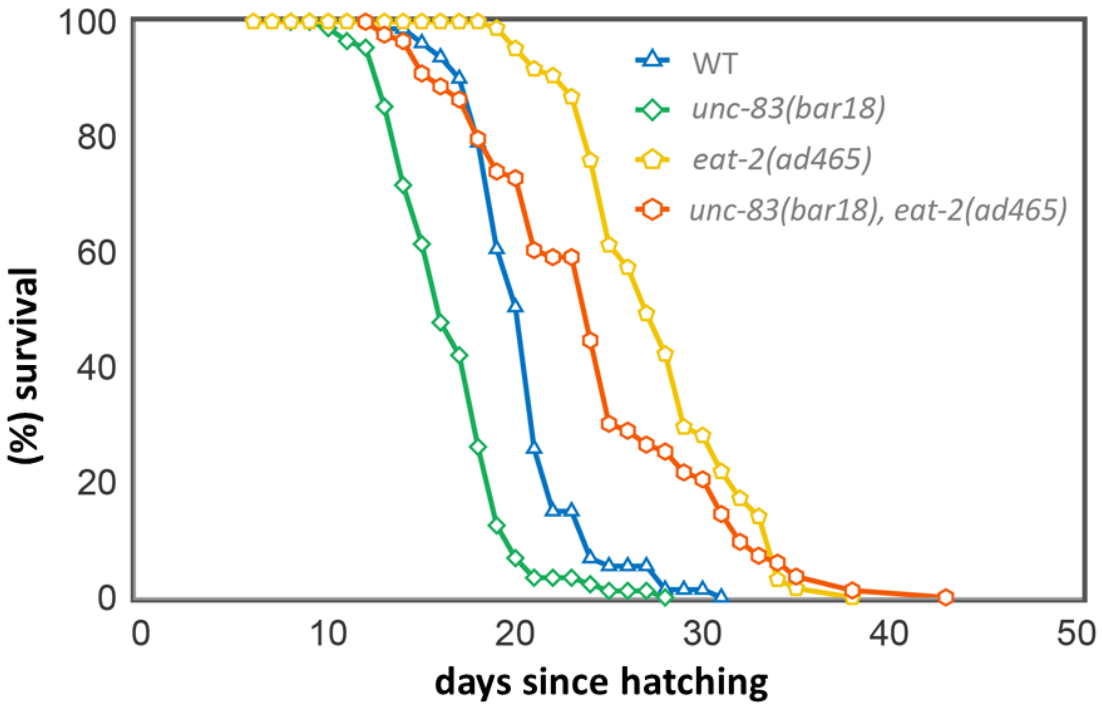
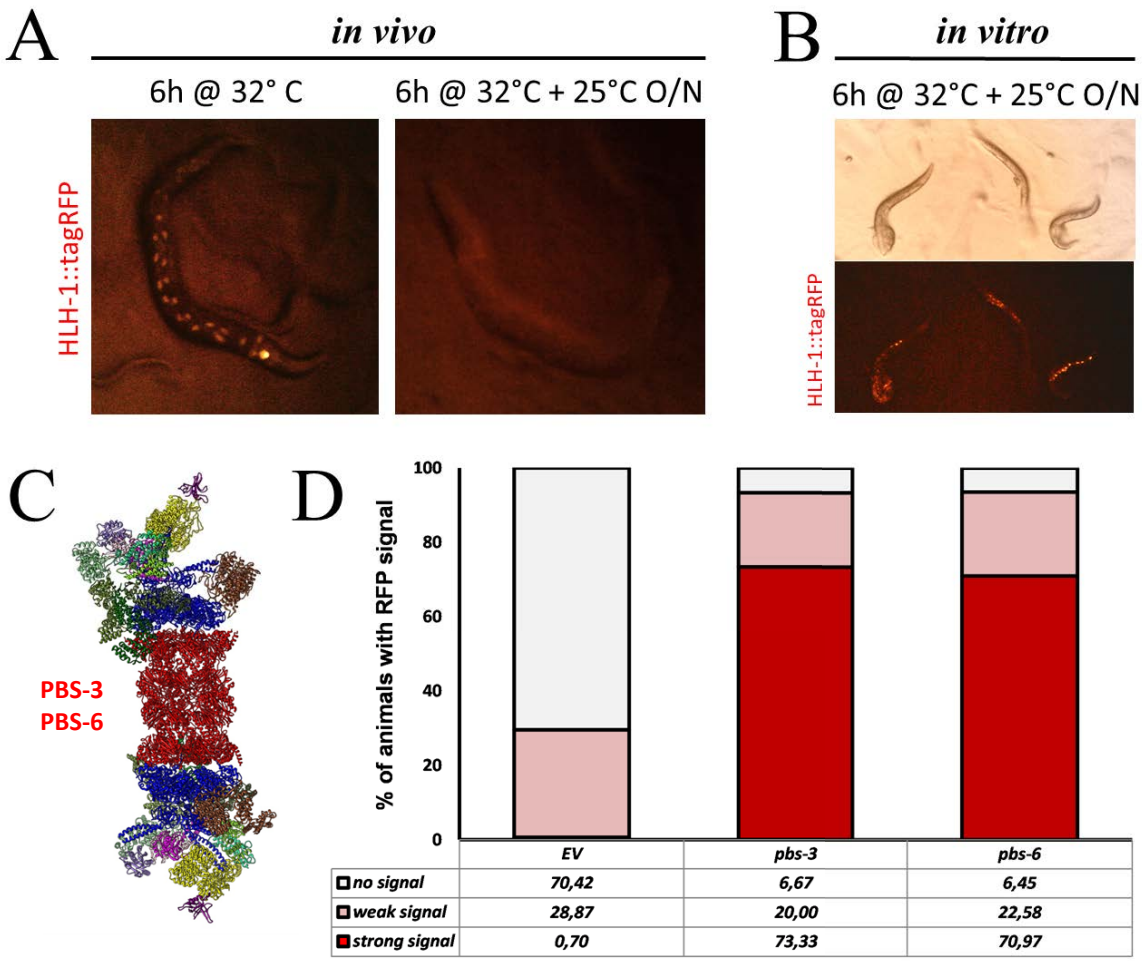


Figure 28. The life extending mutation *eat-2(hx564)* increases the life span of UNC-83 deficient animals to a similar amount, as it does in WT animals. Having the *eat-2(ad465)* mutation in a WT background (yellow line) significantly increases the average life by 35,1% as compared to WT (blue line), while having it in an *unc-83(bar18)* mutant background (orange line) significantly increases the average life span by 45,8% as compared to having the *unc-83(bar18)* mutation only (green line). 90 animals per condition, 3 biological replicates. P-values of all conditions: $p \leq 0{,}01$; Data analyzed using OASIS 2[109].

2.1.8 A HLH-1::tagRFP fusion protein is actively degraded by the proteasome in *C. elegans*

We were wondering why my initial forward genetic screen, which aimed to identify genes that inhibit direct reprogramming of differentiated cells into muscle-like cells, did not yield any such candidates. The only mutant I could isolate showed an accumulation of muscle nuclei around the posterior pharyngeal bulb, a phenotype that turned out to be independent of any *hlh-1* induction. HLH-1 is a TF that specifies muscle cells during embryogenesis and is being maintained in muscle tissue throughout larval development and adulthood. One idea that came to our mind was that maybe HLH-1 protein has a high turn-over rate and needs constant transcription because it's being degraded *in vivo*. To test this hypothesis, I generated a translational reporter line that expresses a HLH-1::tagRFP fusion protein under the control of a heat-shock promoter. Upon heat-shocking animals for 6 h at 32 °C, the protein is clearly visible under a fluorescent microscope, but the signal is lost after keeping the animals at 25

°C O/N (Figure 29A), suggesting that HLH-1::tagRFP proteins get degraded. I repeated the experiment, but killed the worms after the 6 h heat-shock and before shifting them to 25 °C O/N. The tagRFP signal was still visible the next morning, indicating that the protein gets actively degraded *in vivo*, but does not degrade *in vitro* (Figure 29B). One canonical pathway of protein degradation is working through the ubiquitin-proteasome system[112], so we looked up the *C. elegans* genes that would encode for proteins that make up the catalytic center of the 16s proteasome subunit (*pbs-3* and *pbs-6*, Figure 29C).



	EV	pbs-3	pbs-6
no signal	70,42	6,67	6,45
weak signal	28,87	20,00	22,58
strong signal	0,70	73,33	70,97

Figure 29. HLH-1::tagRFP is actively degraded by the ubiquitin-proteasome system. (A) HLH-1::tagRFP translational reporter expression right after a 6 h heat-shock treatment at 32 °C (left) and after an additional O/N incubation at 25 °C (right). (B) Light microscopic (top) and fluorescent (bottom) images of three *C. elegans hlh-1::tagRFP* reporter corpses. Worms were killed after a 6 h heat-shock treatment at 32 °C. Pictures were taken after an additional O/N incubation at 25 °C. (C) 16s Proteasome subunit. The *C. elegans* homologue genes encoding for the catalytic center (red) are shown. Image adapted from Beck et al.[113] (D) Quantification of HLH-1::tagRFP fluorescent signals of the reporter strain shown in A and B, after heat-shock treatment and O/N incubation as in A and B, treated with EV control RNAi, *pbs-3* RNAi and *pbs-6* RNAi. n ≥ 60.

I repeated the experiment another time, treating worms either with a control RNAi (empty vector, EV) or with RNAi against the proteasome genes *pbs-3* or *pbs-6* (Figure 29D). Knocking down either *pbs-3* or *pbs-6* could preserve the HLH-1::tagRFP signal, arguing that HLH-1 gets actively degraded by the ubiquitin-proteasome system.

This fact might be the reason why I didn't identify any reprogramming barriers in my forward genetic screen, as HLH-1, the TF needed to specify muscle fate, seems to get degraded too rapidly. In order to screen for reprogramming barriers that would prevent induced transdifferentiation into a muscle lineage, it might be needed to use a promoter that would ensure constant expression of *hlh-1* in later developmental stages, but would not be active in early embryos, in favor of a heat-shock promoter.

2.2 A reverse genetic screen identifies members of FACT and the pseudogene *F55A3.7* as barriers of induced germ cell to neuron conversion in *C. elegans*

As already outlined in the introduction (section 1.5.1), we identified the FACT complex in *C. elegans* as a reprogramming barrier for the transdifferentiation of germ and intestinal cells to neurons[1]. The human orthologues of the two FACT subunits are SUPTH16 (SPT-16 in *C. elegans*) and SSRP1 (HMG-3 or HMG-4 in *C. elegans*; Figure 30A). Interestingly, SSRP1 has two different orthologues, HMG-3 and HMG-4, which originated from a gene duplication. We were interested to explore if they would have distinct functions in *C. elegans*. An RNAi knock-down of *hmg-3* enables germ cells to be reprogrammed to neurons, while an *hmg-4* knock-down enables intestinal cells to be reprogrammed to neurons, upon broad *che-1* induction. Due to this fact, we thought about the possibility that *hmg-3* and *hmg-4* might be expressed in a tissue specific manner. While for SPT-16, there was an antibody available, no specific antibodies for HMG-3 or HMG-4 existed. Therefore, I HA-tagged both of these genes using CRISPR/Cas9, and indeed, they are differentially expressed. HMG-3::HA could only be detected in the germline (Figure 30B), HMG-4::HA could predominantly be detected in the soma (Figure 30C), while SPT-16 is ubiquitously expressed (Figure 30D). We therefore concluded that SPT-16/HMG-3 FACT safeguards cell identity in the germline, while SPT-16/HMG-4 safeguards cell identity in the intestine (Figure 30E).

While the existence of the two FACT subunits HMG-3 and HMG-4 is the consequence of a gene duplication resulting in two functional proteins with distinct expression patterns, a gene duplication of the FACT subunit SPT-16 resulted in the pseudogene *F55A3.7.* It consists of four exons and is shorter than its mother gene *spt-16.* It is oriented in the opposite direction of *spt-16* less than 2 kb away. A deletion allele spanning the second and third exon is also available (Figure 16A). We found that upon broad CHE-1 induction, mutant animals with the deletion allele *F55A3.7(ok1829)* show a similar germ cell to neuron reprogramming phenotype as upon FACT depletion based on *hmg-3* RNAi (section 1.4.3). Morphological changes like the shape of converted germ cell nuclei and the axo-dendridic like projections of *gcy-5p::gfp* positive cells suggest a robust reprogramming (Figure 17). To our knowledge, this is the first example of a pseudogene whose depletion leads to the permissiveness of a certain tissue to be reprogrammed when challenged by a terminal selector TF and thus I characterized the pseudogene *F55A3.7* and tried to find a potential mechanism for how *F55A3.7* safeguards germline identity.

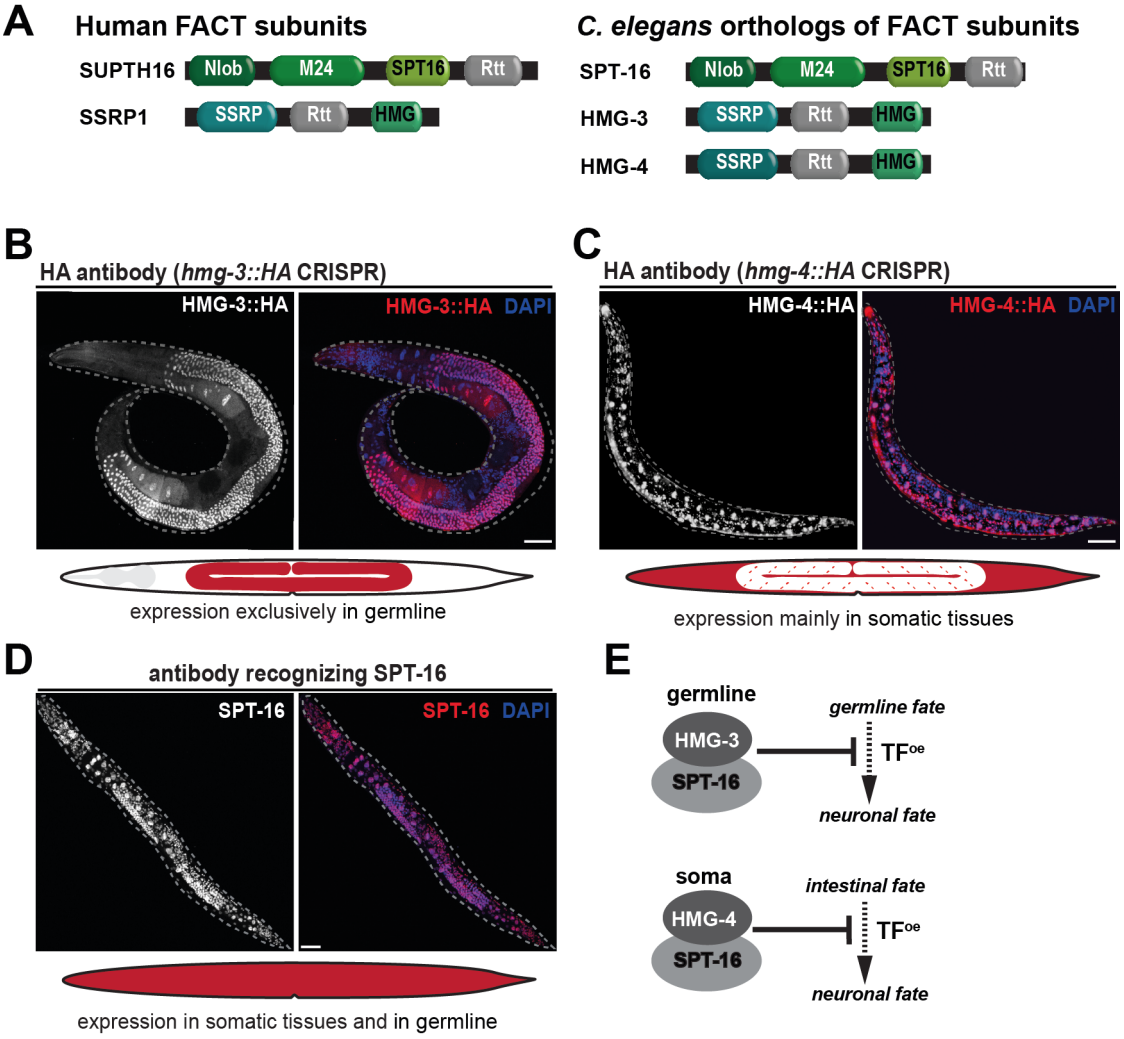


Figure 30. Subunits and expression patterns of FACT in *C. elegans*. (A) Models of FACT subunits in *H. sapiens* and *C. elegans*. Conserved protein domains according to Pfam (pfam.xfam.org) and InterPro (ebi.ac.uk/interpro) are indicated. (B) Immunostaining of CRISPR-tagged HMG-3::HA. (C) Immunostaining of CRISPR-tagged HMG-4::HA. (D) Immunostaining of SPT-16. (E) Model of how tissue-specific FACT isoforms prevent induction of neuronal genes in the germline and intestine. Dashed lines indicate the outline of the animals. Scale bars represent 20 µm.

2.2.1 The pseudogene *F55A3.7* is being transcribed and spliced

To test whether there is a *F55A3.7* transcript being made in the *C. elegans* germline, I dissected gonads from both WT (BAT1753) and *F55A3.7(ok1829)* worms (BAT1749), isolated RNA and reverse transcribed it into cDNA. I designed primers that I could also use for qRT experiments (section 2.2.3), not only for *F55A3.7*, but also for *spt-16* and *hmg-3*.

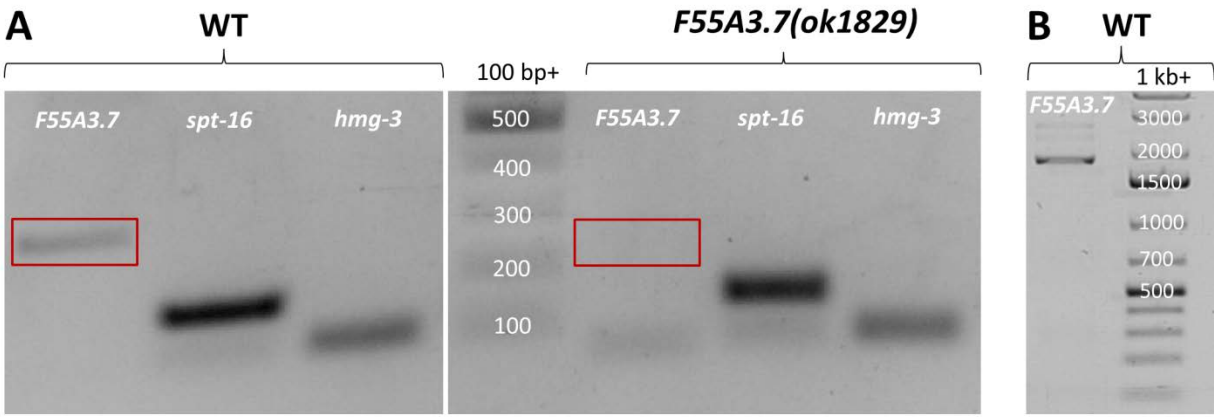


Figure 31. *F55A3.7* is being transcribed and spliced. (A) RT-PCR using WT gonadal (left) or *F55A3.7(ok1829)* mutant (right) RNA as template. Tested genes are indicated. Red boxes mark the *F55A3.7* band or the lack thereof. (B) We amplified the estimated four exons containing full length cDNA transcript (1708bp), which was subsequently subjected to Sanger sequencing.

All primer pairs give a band of the expected size, and there was no band for *F55A3.7* in the *F55A3.7(ok1829)* mutant (Figure 31A). Primers for *F55A3.7* span parts of the third and fourth exon. Additionally, we amplified the estimated full-length cDNA transcript and sent it to Sanger sequencing, thus confirming our predicted intron/exon borders (Figure 31B).

2.2.2 *F55A3.7* RNA is not translated into a protein

To explore the possibility that despite a proper open reading frame (ORF), a small amount of functional protein might be produced from *F55A3.7* RNA due to translational read-through (section 1.5), I tagged the endogenous *F55A3.7* locus with a FLAG epitope using CRISPR/Cas9 editing (Figure 32A). Immunofluorescent staining using an anti-FLAG antibody did not show any signal, unlike the positive control, where I used a strain that expressed LMN-1::FLAG with a nuclear localization signal under the control of the muscle-specific reporter *myo-3p* (Figure 32B).

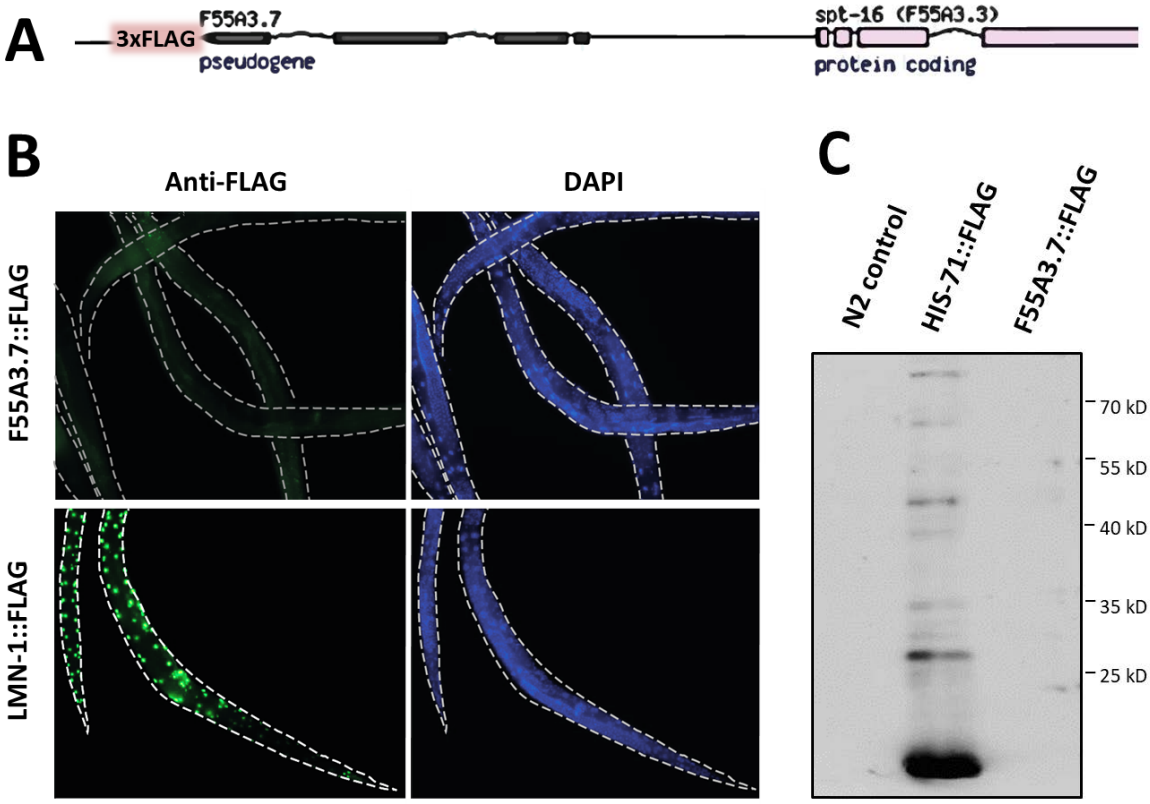


Figure 32. Immunostaining and Western blot of *F55A3.7::FLAG* worm lysates. (A) Using CRISPR/Cas9 gene-editing, a 3xFLAG-tag was introduced at the C-terminal end of endogenous *F55A3.7.* (B) F55A3.7::3xFLAG worm lysates did not show any FLAG staining (BAT1594, upper left), while the LMN-1::2xFLAG positive control did (BAT1468, lower left). DAPI images to the right. (C) Western blot with N2 (negative control), HIS-71::2xFLAG (BAT748, positive control) and F55A3.7::3xFLAG worm lysates stained for FLAG. A positive band for F55A3.7::3xFLAG would have been expected at 24,3 kDa.

Western blot experiments confirmed my results from the immunostainings, as F55A3.7 showed no band (it would have been expected at ~24 kDa), while the positive control for FLAG tagged HIS-71 showed the expected band at ~17 kDa (Figure 32C).

2.2.3 *F55A3.7* ncRNA does not affect mRNA and protein levels of its mother gene *spt-16*, nor those of *hmg-3* or *lin-53*

The simplest explanation for the germline conversion phenotype that we experienced in the *F55A3.7(ok1829)* mutant background (Figure 17) upon broad CHE-1 expression, would be a direct influence of *F55A3.7* ncRNA on transcript or protein levels of its mother gene *spt-16*. To explore this possibility, I performed qRT-PCR of dissected gonads to check for germline-specific transcript levels of *spt-16* in a WT and *F55A3.7(ok1829)* mutant background, and additionally also for the FACT member *hmg-3* and the histone chaperone *lin-53,* which are

both also known to cause the same germline conversion phenotype upon knock-down (section 1.4). However, germ-line specific transcript levels of *spt-16*, *hmg-3* and *lin-53* were unaltered in *F55A3.7(ok1829)* mutant animals (Figure 33A).

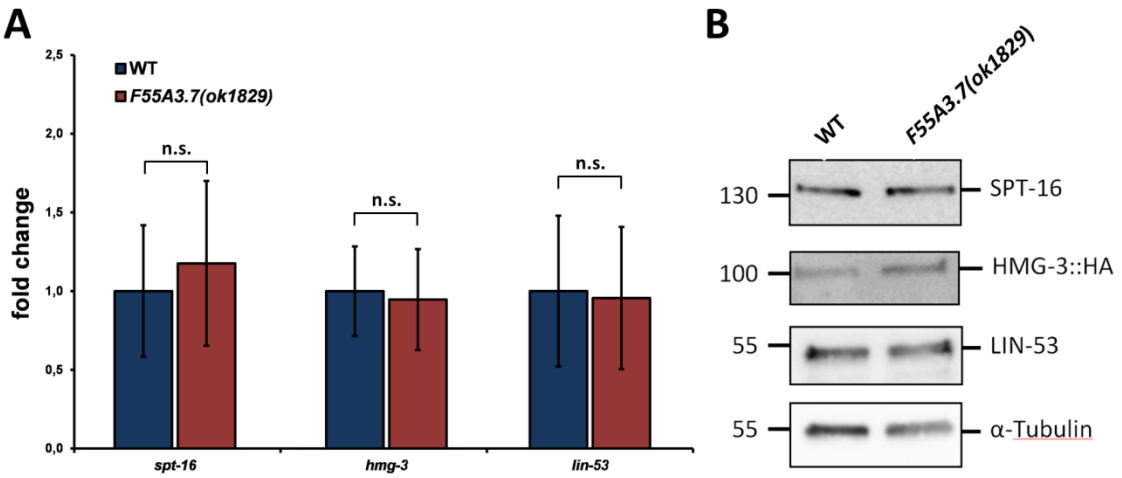


Figure 33. Gonad specific transcript and whole worm protein levels of WT and *F55A3.7(ok1829)* mutant animals for *spt-16*, *hmg-3* and *lin-53*. (A) Gonadal lysates of WT (BAT1753) and *F55A3.7(ok1829)* mutant (BAT1749) animals tested with qRT-PCR for transcript levels of *spt-16*, *hmg-3* and *lin-53*. 3 biological repeats with 3 technical repeats each. Error bars represent SD. n.s.: not significant (B) Western blot of whole worm lysates of WT (BAT1753) and *F55A3.7(ok1829)* mutant (BAT1749) animals. α -Tubulin was stained as loading control after stripping the membrane.

My qRT-PCR results could be supported by western blots, with the only difference that staged young adults were lysed instead of dissected gonads due to higher need of input material. We could show that SPT-16, HMG-3::HA and LIN-53 protein levels were the same in WT and *F55A3.7(ok1829)* mutant animals (Figure 33B), suggesting that the ncRNA *F55A3.7* is acting through other mechanisms to prevent the permissiveness of germ-cells to be reprogrammed into neurons upon *che-1* induction.

2.2.4 *F55A3.7* ncRNA acts in trans

Given the close vicinity of the *F55A3.7* locus to its mother gene *spt-16* (Figure 19A), we wanted to exclude that the germ cell to neuron conversion phenotype caused by *che-1* overexpression in conjunction with the *F55A3.7(ok1829)* deletion allele was an effect in *cis,* acting on *spt-16*. Our qRT-PCR and Western blot data already argue against this possibility, but we still also wanted to make sure that active transcription of *F55A3.7* is needed for it to fulfill its function and that the phenotype we observe is not due to an enhancer element that

got deleted in *ok1829*. In order to test this, I created three recue lines that express WT *F55A3.7* from an extrachromosomal array in the *F55A3.7(ok1829)* mutant background. The strains also have the heat-shock inducible *che-1* construct and the *gcy-5p::gfp* neuronal reporter. After inducing *che-1* by heat-shock in L4 animals, I scored them the next day for the germ line conversion phenotype (Figure 34). For all three lines the phenotype penetrance was reduced to less than 50% of the phenotype penetrance of the *F55A3.7(ok1829)* control strain without any recue construct. This partial rescue demonstrates that *F55A3.7* ncRNA acts in *trans*, as it represses the phenotype even when being transcribed from a different locus. Taken together, our data suggests that *F55A3.7* is being transcribed but not translated and that it acts as an ncRNA in trans, independently of *spt-16*, *hmg-3* or *lin-53*.

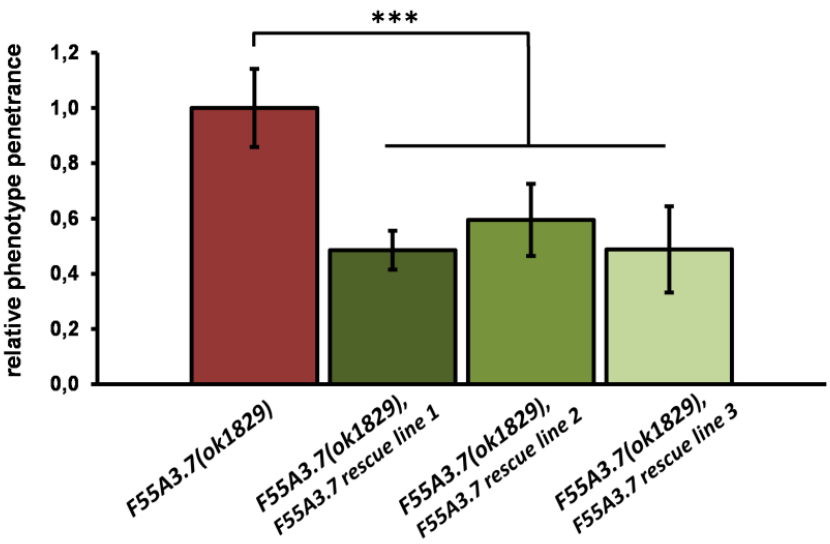


Figure 34. Relative phenotype penetrance of CHE-1 induced germ cells to neurons reprogramming in *F55A3.7(ok1829)* mutant strains with and without a recue construct. Red bar: control strain without rescue construct (BAT372). Green bars: recue strains with rescue construct (BAT1956, BAT1957, BAT1959). n = 44 – 124 per condition; ***: p<0,01

2.2.5 Gonad specific RNA-seq and ATAC-seq reveal many up- and downregulated loci

In order to identify a potential mechanism of how the non-coding RNA *F55A3.7* safeguards germ cell identity, we performed germline-specific RNA sequencing (RNA-seq; Figure 35) and a germline-specific Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq; Figure 36). To do so, we manually dissected gonads and isolated RNA and DNA samples that were subjected to library preparation and sequencing.

RNA-seq uses next-generation sequencing (NGS) to both qualitatively and quantitatively identify RNA in a biological sample. After library preparation (Figure 35A), which includes transcribing RNAs into cDNAs and barcoding samples specifically, the transcriptome of WT

(N2) and *F55A3.7(ok1829)* (RB1524) gonads were sequenced using a NGS Illumina sequencer. Samples were done in triplicates, using 20 gonads per sample. One replicate of *F55A3.7(ok1829)* needed to be dropped, since it didn't pass our quality control due to low sequencing coverage. For the other samples we got, on average, between six and eight million reads per replicate (Figure 35B). N2 control replicates cluster together in a principle component analysis (PCA), as do the *F55A3.7(ok1829)* replicates (Figure 35C) and we could identify several significantly differentially expressed genes between these two groups (Figure 35D).

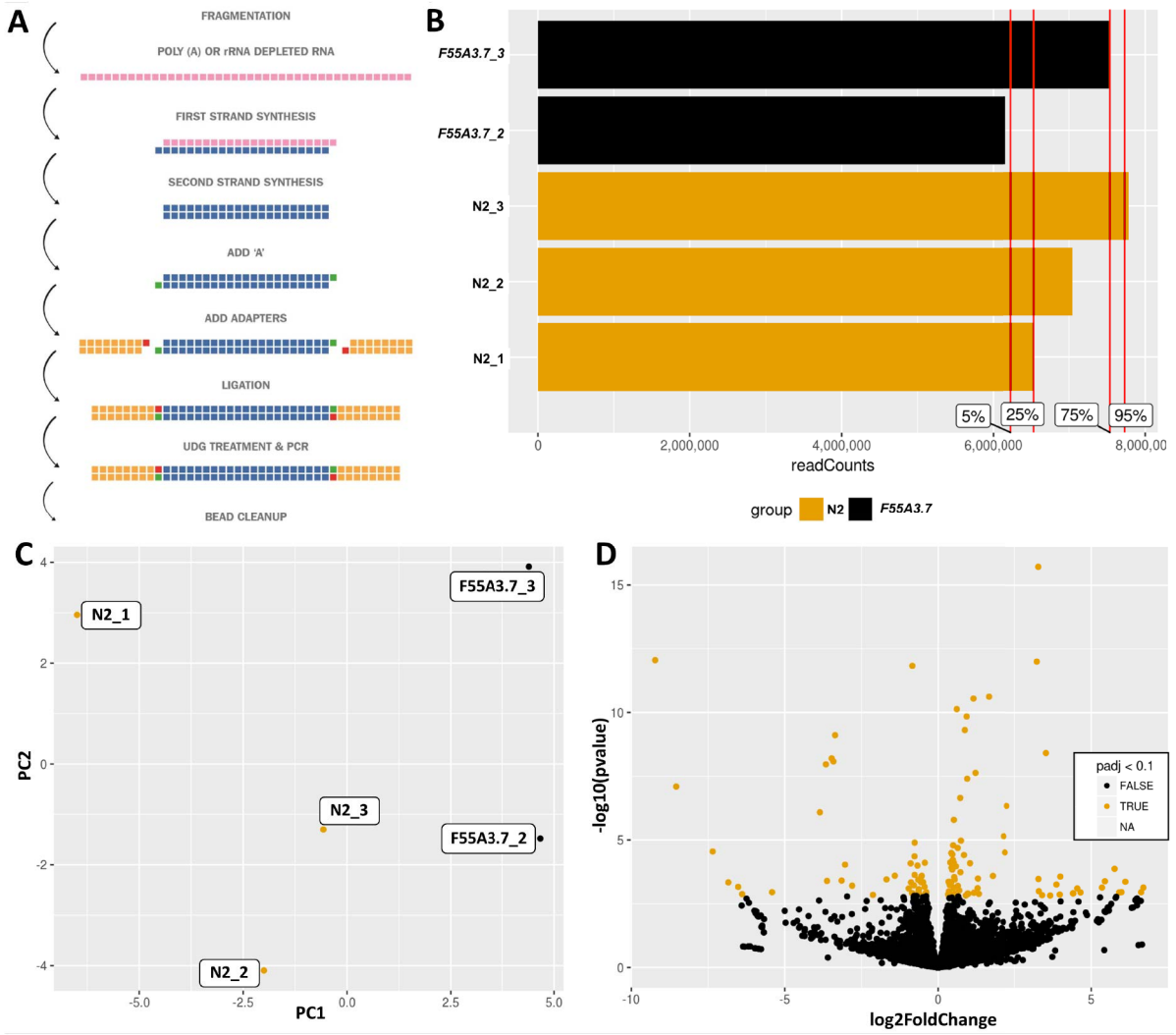


Figure 35. Gonad specific RNA-seq of WT (N2) and *F55A3.7(ok1829)* (RB1524) animals. (A) Library preparation included fragmentation, rRNA depletion, reverse transcription / first strand synthesis, second strand synthesis, A-tailing, adapter ligation and PCR amplification (see material and methods for details). (B) We generated between 6 and 8 million reads per replicate. (C) PCA analysis of WT (N2) vs. *F55A3.7(ok1829)* (RB1524) samples. (D) Differentially expressed genes between WT (N2) and *F55A3.7(ok1829)* (RB1524) gonads.

In addition to RNA-seq, we performed gonad specific ATAC-seq, profiling the epigenetic landscape of DNA extracted from WT (N2) and *F55A3.7(ok1829)* (RB1524) mutant gonads (Figure 36A). The technique makes use of a mutant hyperactive Tn5 transposase, which cuts exposed open DNA and simultaneously ligates specific adapter sequences to it. Adapter-ligated DNA fragments are then isolated, amplified by PCR and used for NGS[114,115]. Triplicates were very similar to each other, but very different between *F55A3.7(ok1829)* sample and N2 control groups (Figure 36B). In general, *F55A3.7(ok1829)* gonadal DNA displayed a more closed chromatin state as compared to N2 DNA, with 4668 loci that were more closed, while only 1063 loci were more open. The loci that were more closed in the *F55A3.7(ok1829)* background were found mostly in promoter regions (97,13%), with the majority of ATAC reads being less than 1 kb from the TSS (86,09%; Figure 36C). In contrast, loci that were more open in the *F55A3.7(ok1829)* background were less likely to be in Promoter regions (74,05%) and also further away from the TSS on average with only 41,03% of loci being within 1 kb of the TSS (Figure 36D).

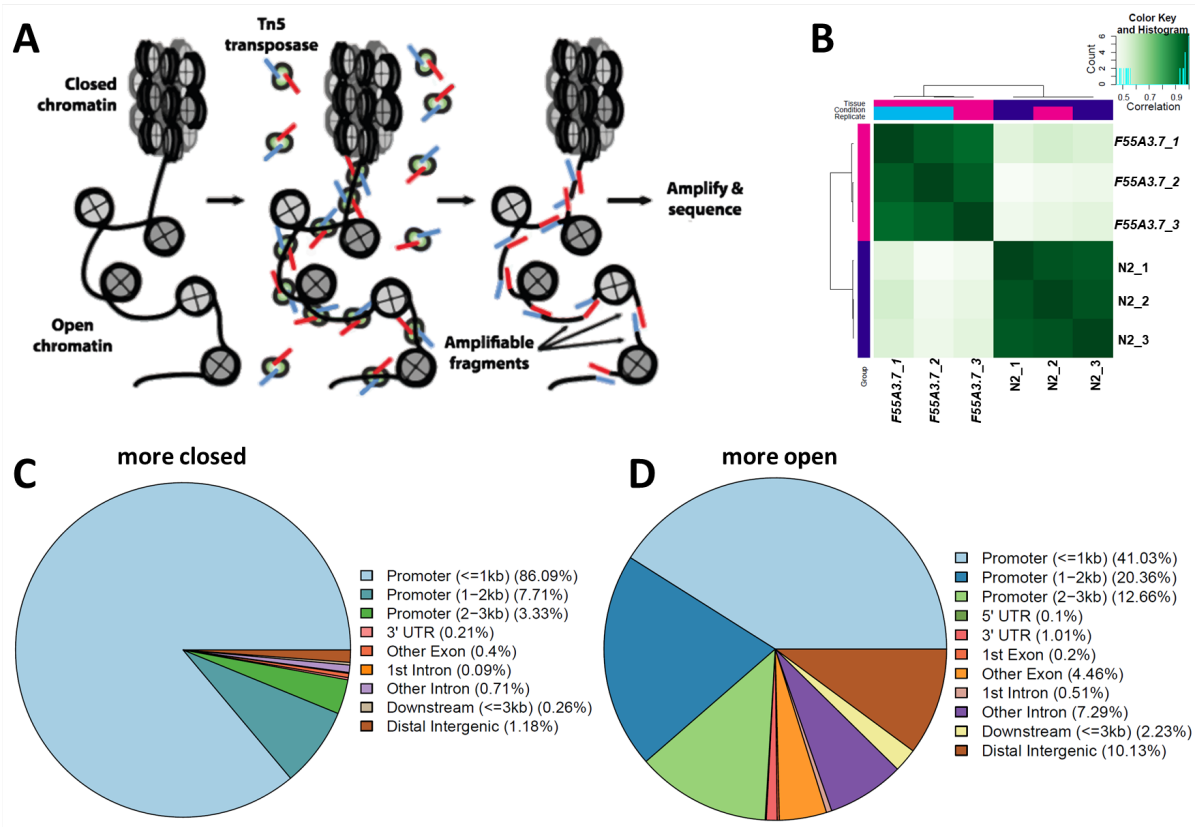


Figure 36 Gonad specific ATAC-seq of WT (N2) and *F55A3.7(ok1829)* (RB1524) animals. (A) A mutant hyperactive Tn5 transposase cuts open DNA regions and simultaneously adds adapter sequences that can be later sequenced with NGS. (B) Heat map of chromatin state similarity between different sequenced samples. (C) Localization of downregulated loci of *F55A3.7(ok1829)* gonadal DNA as compared to WT DNA. (D) Localization of upregulated loci of *F55A3.7(ok1829)* gonadal DNA as compared to WT DNA.

I was looking for overlaps of our RNA-seq and ATAC-seq data sets and compared significantly ($p \leq 0,1$) up-regulated and down-regulated genes that displayed a log2FoldChange ≥ [1,33] (RNA-seq) or a Fold ≥ [1,33] (ATAC-seq). I could find 9 genes that were up-regulated in both data sets (Figure 37A, Table 2) and 12 genes that were down-regulated in both data sets (Figure 37B, Table 2). Furthermore, I compared our data for down-regulated genes with a set of genes that we previously identified as potential reprogramming barriers of induced transdifferentiation into a neuronal lineage by a genome-wide RNAi screen[1]. Doing so, I identified two overlapping genes with our RNA-seq data, including *F55A3.7* itself, which was expected and can be seen as a positive control of our approach. Furthermore, 34 genes were overlapping with our ATAC-seq data (Figure 37C, Table 3). *F55A3.7* down-regulation in the RNA-seq data is shown as an example in Figure 37D.

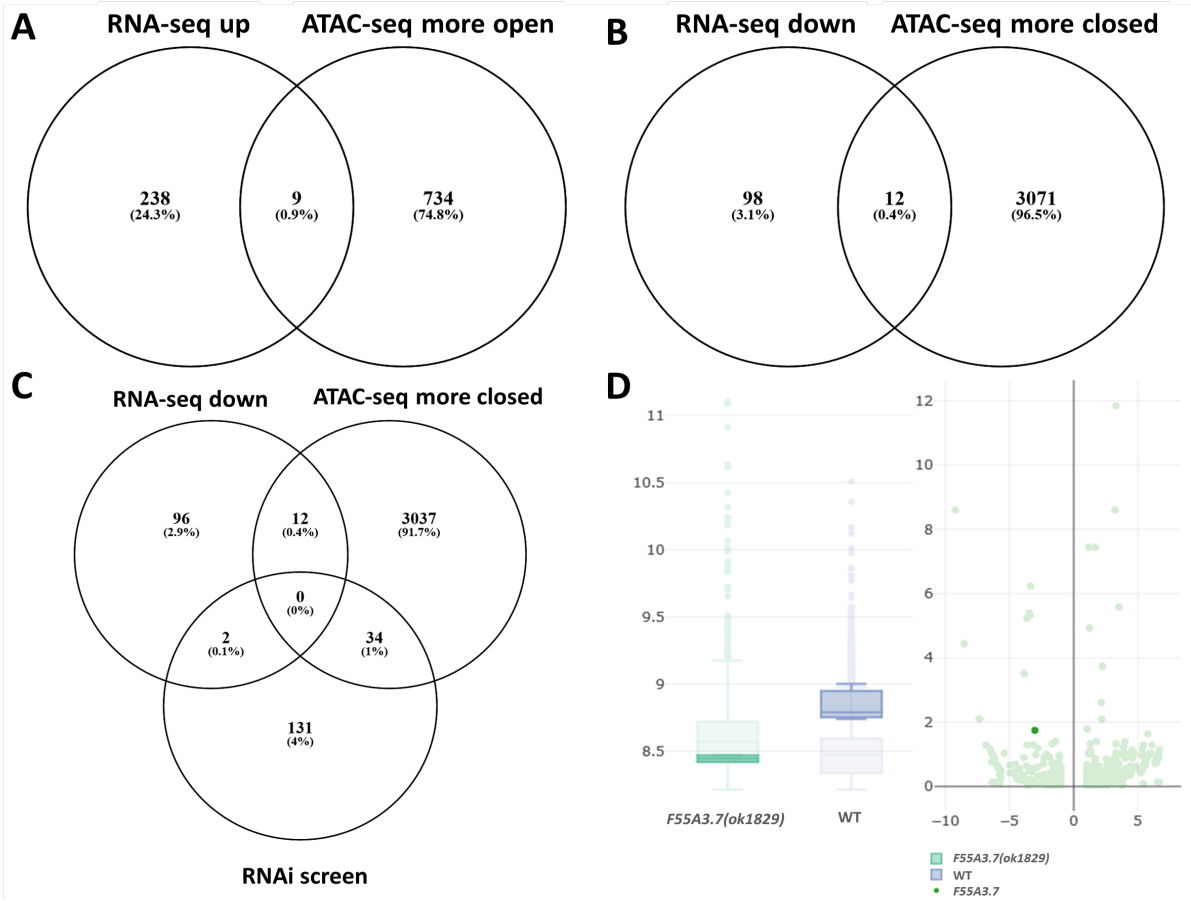


Figure 37. Cross comparisons of RNA-seq and ATAC-seq data of *F55A3.7* depleted worms amongst each other and with a set of previously identified potential reprogramming barriers from a genome-wide RNAi screen. (A) Venn diagram of upregulated genes of *F55A3.7(ok1829)* gonads in comparison to WT gonads for RNA-seq and ATAC-seq. (B) Venn diagram like in A for downregulated genes. (C) Venn diagram like in B, overlapped with potential reprogramming barriers identified by a genome-wide RNAi screen[1]. (D) *F55A3.7* down-regulation in RNA-seq. $P \leq 0,1$ log2FoldChange (RNA-seq) ≥ [1,33]; Fold (ATAC-seq) ≥ [1,33].

Table 2. Overlap of RNA-seq and ATAC-seq data of *F55A3.7* depleted worms.
Upregulated genes colored in green, downregulated ones in red.

Gene Name	Gene biotype	Log2FoldChange (RNA-seq)	p-value (RNA-seq)	Chrom.	Fold (ATAC-seq)	p-value (ATAC-seq)
Y47G6A.31	protein_coding	-4,53	2,74E-02	I	-1,65	5,14E-11
gbb-1	protein_coding	-3,62	4,01E-04	X	-2,15	1,58E-15
fbxc-43	protein_coding	-3,49	1,29E-02	II	-1,63	6,5E-08
his-43	protein_coding	-3,36	7,77E-10	II	-1,66	0,0303
fmo-5	protein_coding	-3,13	1,04E-02	V	-1,8	3,13E-14
unc-51	protein_coding	-2,86	3,75E-02	V	-2,54	6,84E-18
unc-63	protein_coding	-2,43	9,65E-02	I	-1,78	8,1E-09
K07F5.12	protein_coding	-2,25	2,17E-02	IV	-1,76	3,91E-11
fbxa-217	pseudogene	-2,16	3,65E-02	I	-1,75	1,89E-09
CE7X_3.1	pseudogene	-2,15	2,92E-02	X	-1,55	2,22E-07
haf-6	protein_coding	-1,68	3,50E-04	I	-2,28	3,33E-19
bath-19	protein_coding	-1,34	4,62E-02	II	-1,53	6,96E-07
sydn-1	protein_coding	1,41	4,87E-02	I	3,38	2,39E-14
Y48G1C.5	protein_coding	2,92	8,95E-02	I	1,68	8,54E-08
abt-5	protein_coding	3,00	7,03E-02	I	2,95	8,22E-12
best-19	protein_coding	3,36	3,46E-03	IV	1,68	1,25E-09
Y32F6A.6	pseudogene	3,36	8,41E-02	V	1,96	6,23E-11
M162.7	protein_coding	3,40	9,09E-02	V	4,85	4,21E-55
Y43F8C.5	protein_coding	3,57	3,23E-02	V	2,93	1,25E-22
mlc-7	protein_coding	3,74	4,26E-02	III	1,59	5,24E-09
B0545.4	protein_coding	5,18	1,47E-02	IV	3,22	5,52E-22

Together, our sets of overlapping genes (Figure 37A-C) will be the basis of an enhancer/suppressor screen, where I will try to identify genes that upon knock-down of the respective candidate in the *F55A3.7(ok1829)* mutant background, would enhance or suppress the reprogramming efficiency of a germ cell to neuron conversion upon induction of CHE-1. This strategy might help to dissect the molecular mechanism of how *F55A3.7* RNA protects germline identity in *C. elegans*.

Table 3. Overlap of RNA-seq and ATAC-seq data of *F55A3.7* depleted worms with potential reprogramming barriers from a genome-wide RNAi screen[1].

gcy-5p::gfp induction of the genome wide RNA screen depicted in %, affected tissues indicated by abbreviations: GeCo: Germline Conversion, GuCo: Gut Conversion, MuCo: Muscle Conversion, EpCo: Epidermis Conversion.

Gene Name	Gene biotype	*gcy-5p::gfp* induction					Chrom.	Fold (ATAC-seq)	p-value (ATAC-seq)
		GeCo	GuCo	MuCo	EpCo	Other			
cox-5A	protein_coding	0	50	0	0	0	III	-2,77	2,26E-24
Y39B6A.3	protein_coding	0	10	0	0	0	V	-2,74	3,14E-28
afd-1	protein_coding	10	0	10	0	0	I	-2,43	2,81E-17
imb-3	protein_coding	10	0	0	0	0	I	-2,42	7,06E-16
eel-1	protein_coding	10	0	0	0	0	IV	-2,36	1,36E-15
mrg-1	protein_coding	30	0	0	0	0	III	-2,28	2,62E-14
ogdh-1	protein_coding	10	30	0	10	0	IV	-2,15	4,76E-18
lig-1	protein_coding	15	0	0	0	0	V	-2,14	2,35E-11
ucr-1	protein_coding	0	25	0	0	0	III	-2,13	1,56E-15
cdk-9	protein_coding	5	30	0	0	0	I	-2,1	3,00E-14
lpd-5	protein_coding	0	10	0	0	0	I	-2,02	3,38E-16
gfm-1	protein_coding	0	30	0	0	0	II	-2,01	1,88E-15
F23H11.5	protein_coding	0	35	0	0	0	III	-1,97	4,44E-13
R119.5	protein_coding	15	0	0	0	0	I	-1,92	1,04E-09
vha-1	protein_coding	10	0	0	10	0	III	-1,87	7,95E-11
C11H1.3	protein_coding	15	0	0	0	0	X	-1,86	1,80E-07
C34C12.8	protein_coding	0	40	0	0	0	III	-1,77	1,39E-11
C07F11.2	protein_coding	5	0	0	0	0	I	-1,72	4,98E-10
fah-1	protein_coding	0	60	0	0	0	X	-1,66	1,12E-06
dlst-1	protein_coding	0	20	0	0	0	V	-1,65	4,72E-09
F38A1.8	protein_coding	0	50	0	0	0	IV	-1,61	6,58E-10
bud-31	protein_coding	10	0	0	0	0	III	-1,6	2,53E-08
spg-7	protein_coding	30	30	0	0	0	I	-1,59	2,61E-07
Y54E5A.5	protein_coding	15	10	0	0	0	I	-1,58	8,43E-09
cox-5B	protein_coding	15	15	0	0	0	I	-1,57	3,06E-07
div-1	protein_coding	15	0	0	0	0	III	-1,57	1,53E-08
Y82E9BR.3	protein_coding	0	50	0	0	0	III	-1,57	1,98E-07
act-4	protein_coding	15	0	0	0	10	X	-1,53	5,43E-07
hmg-4	protein_coding	0	60	0	0	0	III	-1,49	5,87E-07
tba-2	protein_coding	10	0	0	0	0	I	-1,48	6,79E-07
psf-3	protein_coding	10	0	0	0	0	I	-1,48	4,57E-06
pafo-1	protein_coding	0	40	0	0	0	V	-1,45	3,97E-08
dnj-11	protein_coding	15	0	0	0	0	IV	-1,41	1,56E-06
Gene Name	**Gene biotype**	***gcy-5p::gfp* induction**					**Chrom.**	**Log2Fold Change (RNA-seq)**	**p-value (RNA-seq)**
		GeCo	GuCo	MuCo	EpCo	Other			
his-12	protein_coding	0	20	0	0	0	II	-2,80	6,19E-04
F55A3.7	pseudogene	0	60	0	0	0	I	-3,04	9,28E-05

3 Discussion

In a semi-automated high-throughput forward genetic screen I isolated the mutant *bar18*, showing an accumulation of muscle cell nuclei around the posterior pharyngeal bulb of the worm due to a positioning defect. WGS revealed that *bar18* animals carry a mutation in the KASH-domain gene *unc-83* causing a premature STOP. An additional *unc-83* mutant allele recapitulates the phenotype, as does a mutant allele of UNC-84, a SUN-domain containing protein that interacts with UNC-83. UNC-83 and UNC-84 belong to a LINC complex that bridges the nuclear lamina with the cytoskeleton. Additionally, I could identify several other phenotypes that are connected to a lack of the UNC-83/UNC-84 LINC complex, which have not been described so far: a shorter life span and an impaired exploration and swimming behavior.

Beside developing this screen and analyzing *unc-83(bar18)*, I was also involved in characterizing the FACT complex, which was identified through a whole-genome RNAi screen[1]. Interestingly, one of the FACT complex members, *spt-16*, is the parental gene of a previously uncharacterized pseudogene named *F55A3.7*. A putative null mutant of *F55A3.7,* combined with broad overexpression of CHE-1, showed a germ cells to neurons transdifferentiation phenotype similar to the phenotype observed after depleting the FACT member *hmg-3*.

In this chapter I am going to discuss the differences between forward and reverse genetics, based also on the two screens that I am describing in this thesis. Furthermore, naturally, the results and potential future steps will be discussed.

3.1 Forward vs. reverse genetic screens

As already outlined in the introduction, *C. elegans* is highly amenable to genetic screens due to its short generation time, its hermaphroditic lifestyle and ease of use for both mutagenesis and RNAi knock-down experiments. While my colleague Ena Kolundžić did a reverse genetics genome wide RNAi screen and identified more than 150 reprogramming barrier candidates[1,104], I did a forward genetic mutagenesis screen and identified only one candidate, that later turned out to be no reprogramming barrier (section 2.1, 3.2)[10]. So is reverse genetics superior to forward genetics? At a first glance, it might look like this is the case, but it clearly

isn't, as I will outline in this section. Both approaches have their benefits as well as their drawbacks and, ideally, can complement each other.

The primary reason why I didn't identify more candidates with my semi-automated high-throughput EMS mutagenesis screen might have been the system I used: a transgenic *C. elegans* line allowing ectopic expression of the myogenic bHLH TF MyoD homologue *hlh-1* in combination with a reporter for muscle fate, the transcriptional myosin reporter *myo-3p::gfp::NLS*. The transgenic strain for the genome-wide RNAi screen allowed ectopic expression of the Zn-finger TF CHE-1, which is essential for terminal differentiation of glutamatergic ASE neurons in conjunction with the transcriptional *gcy-5* neuronal reporter. Both strains are making use of the same heat-shock promoter (*hsp-16.2p*) to activate the respective TF. Unlike the strain for the genome-wide RNAi screen, which was already tested and well characterized before[83], the strain I was using was untested. I only found out after having established the semi-automated high-throughput screen that HLH-1 is unstable *in vivo* and actively being degraded by the proteasome system (section 2.1.8). Due to the high turnover of HLH-1 protein, an initial burst of ectopic HLH-1 by heat-shocking the worms might not have been enough to trigger successful transdifferentiation, even when reprogramming barriers were successfully mutated by EMS. To circumvent this shortcoming of the system, an alternative would have been to use a reporter that would be constitutively active during later developmental stages to ensure a constant and stable expression of HLH-1. To our knowledge, such a ubiquitous promoter has not been described yet. Using an early ubiquitous promoter like *eft-3p*[116] would most likely cause several still plastic cells in the early larvae to differentiate to muscle tissue, developmentally arresting the worms or at least rendering them sterile due to the resulting lack of a healthy germline. There are several tissue-specific promoters known that are active only in later developmental stages, like *col-19p*[117] for epidermis or *vit-5p*[118] for gut. This strategy would have the drawback of allowing ectopic expression of HLH-1 in one tissue type only, but would ensure a constant and stable expression of HLH-1. Another possibility would be using a system that allows temporal and spatial control of a gene of interest, like the FLP/FRT and Cre/lox recombination systems[119] or the split cGal system[120].

Another difficulty comes from the fact that the *myo-3* muscle fate reporter necessary for this kind of screen is already very abundantly expressed throughout the worm, highlighting 95 body wall muscle cells among a few others (Figure 6). An increase in the amount of muscle cells can therefore be hard to detect – a reason why I used the Biosorter, which can measure

slight changes in fluorescent patterns more precise than the human eye could. The system used for the genome-wide RNAi screen was based on the ASE neuron fate reporter *gcy-5::gfp,* which very specifically highlights just one ASE neuron (ASER) in the head of the worm. Detecting ectopic GFP signals in other tissues is therefore very fast and easy.

So the discrepancy of the amount of putative reprogramming barriers detected by the forward versus reverse genetic screen, is not due to an intrinsic ability or shortcoming of one of the two techniques, but due to the different setups used to perform the screens (*hlh-1* + *myo-3p::gfp::NLS* vs. *che-1* + *gcy-5p::gfp*). However, what are intrinsic abilities or shortcomings of a forward genetic EMS mutagenesis screen vs. a reverse genetic RNAi screen? Probably the most straightforward advantage of a reverse genetic screen is the fact that a phenotype can immediately be connected to a gene, since the gene being knocked down to induce a certain phenotype is known. In a forward genetic screen, the phenotype-causing gene needs to be identified, which can be done with the help of whole-genome sequencing (WGS). Another advantage of RNAi is the possibility of doing candidate approach screens, testing just a predefined subset of genes. Furthermore, it enables enhancer and suppressor screens, testing a combination of two (or even more) genes, e.g. by using a mutant with a certain phenotype and combining it with RNAi or using conjugation which allows up to three genes to be knocked down simultaneously (Marlon Kazmierzcak et al., unpublished).

The benefits of classical forward genetics by applying chemical mutagenesis are as manifold as the advantages of reversed genetics. Genes can be identified that would show a phenotype only after a complete knock-out and would not show it by applying RNAi, which always results in a partial depletion only. UNC-83, which I identified during a forward genetic EMS screen, is a good example. The allele I isolated contains a point mutation causing a premature STOP, and this putative null-allele causes muscle nuclei to be mis-positioned (section 2.1). Interestingly, we could not recapitulate this phenotype by applying RNAi against *unc-83* in a WT background, but could do so using other premature STOP alleles of *unc-83*, indicating that a complete depletion of UNC-83 is needed to cause this specific phenotype. Beside phenotypes that require a complete knock-out rather than a knock-down of a gene, forward mutagenesis screens can potentially also identify gain of function alleles, e.g. by causing a mutation that increases transcription of a gene or a mutation that would alter post-transcriptional regulation. Finally, mutagenesis screens might also isolate temperature sensitive alleles, alleles where the gene product is stable at a certain temperature, but is rendered nonfunctional at a specific non-permissive temperature. Such alleles enable loss of

function studies of genes that would, for instance, be embryonically lethal, by allowing the shifting of animals at a later developmental stage to the non-permissive temperature, thus enabling a depletion of the mutated gene product.

Taken together, both forward and reverse genetics have several unique advantages and disadvantages. Depending on the scientific question, one may be chosen over the other, or it might even make sense to apply both for a more complementary approach. The semi-automated high-throughput forward genetic screen I developed proofed to be very efficient, isolating a mutant that displayed a rather subtle phenotype, which was based only on some mis-positioned GFP positive nuclei that the BioSorter was able to detect. Combining the screen that I developed with an alternative approach to ectopically express *hlh-1* (as discussed above) should be a very effective way of screening for factors that counteract HLH-1 induced transdifferentiation.

3.2 A forward genetic screen aiming to find reprogramming barriers in *C. elegans* reveals something unexpected

The beauty of forward genetics relies on the fact that it's a rather chaotic approach, so unexpected outcomes can be expected. In my case I isolated the *unc-83(bar18)* allele, which we initially thought was causing an increase in the number of body wall muscle cells, since mutant animals displayed a strong accumulation of body wall muscle cells around the posterior pharyngeal bulb (Figure 19C). Following careful analysis, and quantifying the amount of muscle cells in WT and *unc-83(bar18)* mutant animals, which remained unchanged as compared to WT animals (Figure 21), I realized that this was not the case. Based on this data and due to already published information about the UNC-83/84 LINC complex (section 1.3.2), I realized that the observed phenotype was due to a positioning defect of muscle nuclei, rather than a transdifferentiation event, especially since ectopic expression of the TF *hlh-1* was not necessary to trigger that phenotype. My data suggests that this cell autonomous effect is primarily due to nuclei that are displaced from the head of the worm towards the neck region (Figure 22). An alternative explanation would be that some cells in the head region are missing, while some muscle cells in the neck region undergo additional cell divisions, thereby leading to an accumulation of nuclei (and cells) in the neck area. I think this scenario is rather

unlikely, since it can be explained by a similar nuclear migration mechanism that has been already shown for UNC-83/84 LINC in other lineages (section 1.3.2), but I cannot exclude that additional cell divisions might occur.

3.3 The role of the UNC-83/UNC-84 LINC complex in *C. elegans* body wall muscle cells

Here, I describe a so far uncharacterized role of the UNC-83/UNC-84 LINC complex in positioning body wall muscle nuclei in *C. elegans*. Animals having mutations in *unc-83/unc-84* display an accumulation of body wall muscle nuclei around the posterior pharyngeal bulb. The phenotype is cell autonomous, as *unc-83(bar18)* animals can be rescued by driving WT *unc-83* from a muscle specific *myo-3* promoter. Using a ubiquitous *eft-3* promoter rescues the phenotype as well.

The reporters I used to show the nuclei mis-positioning effect in *unc-83(bar18)* animals are nuclear, so they cannot distinguish between mis-positioning of whole cells or of nuclei only. Since the somatic muscle tissue in *C. elegans* is not syncytial (unlike the epidermis), it's likely that whole cells are mis-positioned, especially since this is also the case in other lineages, such as migrating P cells[76–79]. Using different reporters or immunofluorescent staining could potentially verify this assumption.

What was quite interesting to see as well was the fact that accumulating nuclei around the second pharyngeal bulb originated mostly from the head region and not the posterior body, although the latter accommodates way more nuclei than the head region (Figure 22). It might be that the UNC-83/UNC84 LINC complex is mostly active in those muscle cells and not so much in others. Alternatively, it could also very likely be that the mis-positioning is a result of a general absence of this LINC complex and the phenotype is of passive nature, resulting from physical restraints and arrangements of striated body wall muscles to contribute forming the outer tube of the worm.

One interesting aspect of the role of the UNC-83/UNC-84 LINC complex in muscle cells is the absence of a strong and obvious phenotype connected to it. In contrast, mis-positioned P cells for instance undergo apoptosis, which results in a lack of neurons in the ventral nerve cord (section 1.3.2), as well as impaired backwards movements as a secondary effect thereof. While this is mostly happening at elevated temperatures (25 °C), I could describe additional Unc phenotypes using exploration and swimming assays, which were carried out at 15 °C.

However, the Unc phenotypes seem to be mostly independent from the mis-positioned muscle nuclei phenotype. Both *myo-3p* and *eft-3p* rescue lines fail to rescue the exploration behavioral defect, suggesting that it is not linked to mis-positioned nuclei in muscle cells (since the mis-positioning phenotype is rescued in both lines, see Figure 19C, 20A). This could be due to a copy number issue of the rescue constructs and the exact expression levels of *unc-83* might would need a certain fine tuning to match WT expression levels (there might very likely be an overexpression of UNC-83 protein in the rescue lines). In contrast to the impaired exploration behavior, impaired swimming could be fully rescued by driving WT *unc-83* from the ubiquitous *eft-3* promoter, but not from the muscle specific *myo-3* promoter (Figure 25). This result suggests that that a lack of UNC-83 in a tissue other than muscle is responsible for this phenotype or, alternatively, a combination of muscle and another tissue.

Besides the effect of lacking UNC-83 on motility, I also analyzed the effect it has on *C. elegans'* life span and could detect a shorter life span of UNC-83 deficient animals as compared to WT animals (Figure 26). Interestingly, neither driving WT *unc-83* from the muscle specific promoter *myo-3p* nor from the ubiquitous promoter *eft-3p* could fully rescue the short life span phenotype, but both rescue lines still showed a significant partial rescue (Figure 26). The absence of a full rescue could be a copy number issue as discussed for the motility assays. Furthermore, these results support a model where the lifespan decrease is linked to an absence of the UNC-83/84 LINC complex specifically in muscle tissue, since driving the rescue construct from the ubiquitous *eft-3* promoter doesn't further increase the amount of rescue as compared to using the muscle-specific *myo-3* promoter. The exact mechanism of the lifespan shortening remains to be determined, but there have been studies already that connect LINC complexes to premature aging and disease[121,122]. I generated *unc-83* and *age-1* as well as *unc-83* and *eat-2* double mutants to test in lifespan experiments. My results (Figure 27, 28) support the hypothesis that the effect of lacking UNC-83 on *C. elegans'* life-span is unrelated to both the *daf-2* pathway and to dietary restriction.

Taken together, I could identify several phenotypes that are connected to a lack of the UNC-83/UNC-84 LINC complex, which have not been described so far: mis-positioned body wall muscle nuclei, a shorter life span, impaired exploration and swimming behavior. Also, phenotypes related to muscle tissue have not been described in this context yet (only for P cells, the intestine and hyp7 hypodermal precursors) and it might very likely be that in the future other novel phenotypes will be connected to the UNC-83/UNC-84 LINC complex,

especially since it is ubiquitously expressed throughout the worm and also due to the fact that mis-positioned nuclei might work upstream of, and thereby affect many, other processes.

3.4 The ncRNA *F55A3.7* does not affect its mother gene *spt-16* and acts in trans to protect germline fate

In contrast to the UNC-83/UNC-84 LINC complex, the FACT complex in *C. elegans,* consisting of SPT-16 and either HMG-3 or HMG-4, is not ubiquitously expressed. Using CRISPR/Cas9 gene editing to introduce HA-tags to both HMG-3 and HMG-4 in conjunction with immunostaining, I could show that HMG-3 is exclusively present in the germline and HMG-4 is predominantly present in the soma, confirming our hypothesis that two alternative FACT complexes exist in *C. elegans*: one soma- and one germline-specific (Figure 30). While in the case of *hmg-3* and *hmg-4* a gene duplication resulted in two functional proteins with distinct expression patterns, a gene duplication of the FACT subunit *spt-16* resulted in the non-translated pseudogene *F55A3.7.*

I could show that the non-coding RNA *F55A3.7* does not directly influence its mother gene and FACT member *spt-16*, performing both qRT-PCR and Western blots of *F55A3.7(ok1829)* mutant and control worms (Figure 33). The qRT-PCR experiments were performed with dissected gonads, while for Western blots, staged L4 worms were used. *F55A3.7* ncRNA regulating *spt-16* expression would have been the simplest explanation for the germline conversion phenotype that we experienced in the *F55A3.7(ok1829)* mutant background upon broad CHE-1 expression (Figure 17). However, our data suggests that *F55A3.7* is acting through other mechanisms to block the permissiveness of germ cells to be reprogrammed into neurons. Furthermore, SPT-16 is part of both a somatic and a germline-specific complex and we could previously demonstrate that gut cells can be reprogrammed to neuron-like cells upon RNAi knock-down of *spt-16* and broad CHE-1 expression using RNAi[1,104]. If *spt-16* was to be down-regulated in the *F55A3.7(ok1829)* mutant background, we would have expected to see the same, but germline conversion was the only phenotype we experienced upon broad CHE-1 induction in *F55A3.7(ok1829)* animals. It is worth mentioning, that on the other hand, we could never show germline reprogramming in animals where we knocked-down *spt-16* with RNAi, because *spt-16* F1 RNAi during embryonic development is causing early lethality. F1 RNAi is needed to assay for conversion of germ cells, as shown for the depletion of *hmg-3*

or other previously identified factors such as *lin-53*[83,84] (section 1.4). Interestingly, using RNAi against *F55A3.7* resulted in intestinal cells starting to convert to neurons upon broad CHE-1 induction[1,104], unlike the deletion mutant *F55A3.7(ok1829)* where we could just see germ-cell reprogramming. This counter-intuitive result can be easily explained by the high sequence homology between the pseudogene *F55A3.7* and its mother gene *spt-16*, so *F55A3.7* RNAi also targeted *spt-16* transcripts, resulting in a *spt-16* knock-down phenotype and displaying the same lethality in an F1 screen that has been explained above.
Finally I could demonstrate that *F55A3.7* acts in trans by doing rescue experiments (Figure 34), thus confirming that active transcription of *F55A3.7* is needed for it to fulfill its function and that the transdifferentiation phenotype we observe is not due to an enhancer element that got deleted in the *F55A3.7(ok1829)* deletion allele.

3.5 Gonad specific transcriptome and epigenetic landscape analysis in *F55A3.7* depleted worms

We performed germline-specific RNA-seq and ATAC-seq in order to identify a potential mechanism of how the non-coding RNA *F55A3.7* safeguards germ cell identity (section 2.2.5). While the RNA-seq data has a rather low read depth, based on our experience with NGS of *C. elegans*, the read depth of the ATAC-seq data looks much better and covers many more genes. Despite the fact that we might repeat the RNA-seq experiment with a higher input amount of dissected gonads, we already have several overlapping genes that are high-confidence candidates for an enhancer/suppressor screen.
Furthermore I compared the data for down-regulated genes of RNA-seq and more closed genes of ATAC-seq with a set of genes that we previously identified as potential reprogramming barriers of induced transdifferentiation into a neuronal lineage by a genome-wide RNAi screen[1], which is giving me another set of high-confidence candidates to test. In my ongoing experiments I am at the moment knocking down these candidates in the *F55A3.7(ok1829)* mutant background and check if upon ectopic induction of *che-1* the reprogramming efficiency of a germ cell to neuron conversion would be enhanced or suppressed.

Taken together, *F55A3.7* plays a role is safeguarding cell identity of the germline independently of its mother gene *spt-16* or other FACT members. It therefore must act through another mechanism which still needs to be elucidated. My RNA-seq and ATAC-seq data provided us with lots of clues in this regard and I am currently using this information to conduct enhancer and suppressor RNAi screens.

Generally, reprogramming barriers are often evolutionarily conserved: FACT safeguards germline and intestinal fate in *C. elegans* and also safeguards cell identity of human fibroblasts[1]. LIN-53 (CAF-1p48/RBBP7 in mammals) is another reprogramming barrier that safeguards germline fate in *C. elegans*. Additionally, the LIN-53-containing histone chaperone CAF-1 acts as a barrier of reprogramming murine fibroblasts[83,85]. Naturally, metazoan organisms need to prevent induction of ectopic gene expression programs in order to be successful and therefor such evolutionary conservation is quite likely.
Due to this fact, *C. elegans* can serve as a very versatile and powerful model organism to identify such evolutionary conserved reprogramming barriers or other genes, as also demonstrated in this thesis.

4 Material & Methods

4.1 Nematode culture

C. elegans strains were maintained on nematode growth medium (NGM) seeded with OP50 bacteria at 15 °C, 20 °C or 25 °C as described previously[3]. Temperature was chosen according to the experimental setup. If not mentioned otherwise, the temperature used was 20 °C.

4.2 Transgenesis in *C. elegans*

4.2.1 Microinjection into *C. elegans'* gonads

Using a fine needle, DNA can be directly injected into the germline of *C. elegans*, where it usually forms an extrachromosomal array that is transmitted to some individuals of the F1 offspring[123]. Usually, besides the DNA of interest, a marker is used to distinguish animals that carry an extrachromosomal array from the ones that don't. Typical markers are for instance the phenotypic marker *rol-6* (the plasmid pRF4 encodes *rol-6(su1006),* which is a dominant mutation of a collagen gene that leads to a horseshoe-shaped movement)[124], *myo-2p::mCherry* (expression only in the pharynx of the worm)[125] or the antibiotic resistance gene against Hygromycin[126]. In order to generate extrachromosomal arrays where germline expression was needed, a complex array strategy was used[127]. Using this strategy, the DNA concatamerizes, which prevents the formation of highly repetitive arrays, that are usually silenced in the germ line[128].
Microinjections where done by Sergej Herzog or Alina El-Khalili.

4.2.2 Integration of extrachromosomal arrays with γ-irradiation

Extrachromosomal arrays can be integrated into the genome using γ-irradiation. To do so, 150 L4 larvae or young adults are irradiated at 4.000 rad. Subsequently, worms are transferred to fresh plates with 5 worms each. After the worms propagated and went into starvation (usually 10 - 12 days at 20 °C), plates are chunked and 10 worms of each plate are singled the next day. The next generation is scored for the transmission rate of the extrachromosomal array. Animals that give 100 % positive transgenic F1s are considered to be integrated. Resulting

lines need to be backcrossed to WT animals 3 - 5x to reduce the amount of unwanted mutations caused by the γ-irradiation.

4.2.3 Universal MosSCI

MosSCI is a method to insert a single copy of a transgene into a well-defined location in the *C. elegans* genome (see section 1.2.2.1). Universal MosSCI sites are insertion sites that allow insertion of the same transgene at several well-defined locations with the advantage that one targeting vector is compatible with multiple insertion sites based on the strain that is used for mircoinjection. Universal MosSCI was applied following the instructions of www.wormbuilder.org

4.2.4 Gene tagging using CRISPR/Cas9

FLAG-tagging of *F55A3.7* to generate BAT1594, HA-tagging of *hmg-3* to generate BAT1559 and HA-tagging of *hmg-4* to generate BAT1967 was performed using the CRISPR/Cas system following the instructions from Arribere et al.[129]. The injection mixes contained a PCR repair template to introduce the tag (90ng/µl, ~50bp homology arms), a plasmid that drives Cas9 expression (Addgene #46168) and a plasmid expressing the sgRNA targeting the *F55A3.7* (dBT625, 50ng/µl), *hmg-3* (dBT620, 50ng/µl) or hmg-4 (dBT621, 50ng/µl) locus. Furthermore a *dpy-10* co-CRISPR approach was used[129] to facilitate screening, so the injection mix also contained a *dpy-10(cn64)* PAGE-purified 99mer single-stranded oligodeoxynucleotide (ssODN) HR template (50ng/µL; IDT) and a plasmid that drives expression of dpy-10(cn64) sgRNA (dBT618, 50ng/µl). Successful knock-in was confirmed by Sanger sequencing (Eurofins).

4.3 Transgenic crosses of *C. elegans*

Mating was done by transferring 7 males of one strain and 3 hermaphrodites of another strain to a plate with only limited availability of food. This enforced the worms to stay in close proximity, thus increasing the chance of successful mating. Prior to that, WT (N2) males were crossed into a strain to generate males of this strain if otherwise no males were available. Genotypes were monitored by phenotypical markers and/or by genotyping.

4.4 Identification of the mutated *bar18* locus with WGS

To identify the mutated *bar18* locus, I applied a Hawaiian SNP crossing strategy and combined it with whole-genome sequencing (WGS) as described by Doitsidou et al.[105]. Using this strategy I introduced Hawaiian SNPs through crossing and homologous recombination into the genome of the mutant and then narrowed down the location of the mutation following WGS. After each cross, only animals that still show the mutant phenotype are analyzed and, as a result, fewer or no Hawaiian SNPs will be found around the *bar18* locus, making it possible to identify it.

Briefly, I crossed *bar18* mutant (BAT173) and WT (BAT60) animals with Hawaiian CB4856 males and singled F1 cross progeny. After they self-fertilized and propagated, I singled 41 independent F2s of *bar18* worms that showed the nuclei displacement phenotype and 36 independent F2s of WT worms to fresh plates and allowed them to self-fertilize. After their progeny populated 6 cm NGM plates, worms were washed off with M9 and their DNA was isolated using the Gentra Puregene Tissue Kit (Qiagen, #158689) according to the manufacturer's instructions. Pooled DNA samples of *bar18* and WT worms were used to prepare libraries with the Paired-End Sample Prep Kit (Illumina #PE-102-1002) according to the manufacturer's protocol. Libraries were subjected to WGS on an Illumina HiSeq 2500 System, using single 100 nucleotide reads. Bioinformatic analysis was done using the Galaxy-based CloudMap tool according Minevich et al.[106].

4.5 Microscopic analysis of *bar18* phenotype

To quantify the nuclei dispositioning phenotype penetrance (Fig. 1B), living worms were assessed under a fluorescent dissecting microscope. To count total numbers of GFP-positive nuclei, worms were immobilized with polystyrene microspheres as described by Fang-Yen et al.[130]. I took image stacks of whole worms using a Leica DM6B-Z microscope. Stacks were opened using Fiji (http://fiji.sc) and nuclei quantified with the help of the multi-selection tool to mark counted nuclei. Sex muscles (vulva and uterine muscles) were excluded when counting.

4.6 Lifespan assays

To monitor the lifespan of different strains, synchronized eggs were put on OP50 plates and worms were grown until L4 stage at 15 °C. Subsequently, L4 larvae were transferred to FUDR plates and shifted to 20 °C. FUDR inhibits DNA synthesis and prevents the hatching of eggs and larval growth[131,132], so worms that are monitored for their lifespan are not overgrown by their progeny. The lifespan was scored every day by counting the number of dead vs. alive animals. Data was analyzed using the online application for survival analysis 2 (OASIS 2)[107].

4.7 Swimming assays

Swimming assays are one way to analyze the motility of worm quantitatively. To do so, a single worm was transferred to a droplet of M9 buffer (10 µl) and filmed by a microscope camera. Videos were captured using a light microscope, DinoEye eye piece camera and VirtualDub 1.10.4 software, with the following parameters: 640x480 pixels, 16 seconds, 30 frames per second. The videos were imported into ImageJ software, cutting the first second of each video to circumvent flickering of the first few frames. Subsequently, the videos where analyzed using the ImageJ plugin wrMTrck[111] with the following parameters for the wrMTrck batch: fileTypeNr=1; imageType=1; backSub=0; threshMode=1; fixedThresh=13; skeletonize=0; movieDuration=0; fps=30; pixPrMm=0; minSize=40; maxSize=700; maxVelocity=100; maxAreaChange=500; minTrackLength=20; bendThreshold=2; bendDetect=2; rawData=0; delFirst=0. Finally the body bends per second were calculated by dividing the measured amount of body bends by the time.

4.8 Exploration assays

An exploration assay measures exploration behavior based on movement across a bacterial lawn that has a grid applied to it. They were performed as previously reported[110]. Briefly, individual L4 animals were transferred to a 6 cm agar dish seeded with OP50. After ~16 h, the worm was removed and the plate imposed to grid with 3,5 mm squares and the number of squares entered was manually counted and squares entered per hour was calculated.

4.9 RNAi experiments

RNAi plates were made from NGM-Agar, additionally supplemented with IPTG (1 mM final concentration) and carbenicillin (50 µg/µl final concentration) and seeded with 100 µl overnight culture of RNAi bacteria (HT115). HT115 bacteria contain an IPTG-dependent T7 polymerase to express dsRNA[70] (section 1.2.2.2). Additionally their dsRNA RNaseIII gene is knocked out and substituted with a tetracycline resistance. RNAi bacteria were streaked out on LB-agar plates containing ampicillin or carbenicillin (selection for the RNAi plasmid) and tetracycline (selection for RNaseIII disruption). From those plates, overnight cultures of HT115 were done in LB medium containing only ampicillin, since tetracycline would have influenced RNAi efficiency.
For RNAi experiments, worms were grown on plates seeded with RNAi bacteria from Ahringer library (Source Bioscience). RNAi against Renilla luciferase (Rluc) was used as negative control. Experimental procedures followed a standard RNAi feeding protocol[72].

4.10 Immunostaining of *C. elegans*

For anti-FLAG staining (Anti-FLAG mono mouse antibody, Sigma, at 1:1000 dilution), whole worms were fixed and permeabilized as described previously described method[133]. Briefly, worms were washed 3x and resuspended in RFB (160 mM KCl; 40 mM NaCl; 20 mM EGTA; 10 mM Spermidine) + 2% formaldehyde, followed by three freeze-thaw cycles and an incubation for 30 minutes at 25 °C. After that, samples were washed with TTE (100 mM Tris pH 7,4; 1 % Triton; 1mM EDTA) and incubated for 4 h at 37°C with shaking in TTE + 1% beta-Mercaptoethanol. After that, samples were washed in BO3 buffer (10 mM H3BO3; 10 mM NaOH; 2 % Triton) and incubated for 15 min at 37°C with shaking in BO3 buffer + 10 mM DTT. Following another wash with BO3, samples were incubated in BO3 buffer + 0,3 % H_2O_2 for 15 minutes at 25 °C. After a final wash in BO3, samples were blocked 0,2 % gelatin + 0,25 % Triton in PBS and stained. Fixed worms were incubated in Anti-FLAG diluted primary antibody (1:1000 in PBS with 0,25 % Triton + 0,2 % gelatin) overnight at 4 °C. After washing in PBS + 0,25 % Triton, secondary antibodies (AlexaFluor488 Goat Anti-MouseAlexa, Molecular Probes, at 1:1500 dilution) were applied and incubated overnight at 4°C. Finally, samples were washed in PBS + 0,25 % Triton and mounted on glass slides with DAPI-containing mounting medium (Dianove, #CR-3448).
Anti-HA staining was performed in the exactly same manner, only using another primary antibody (anti-HA mono mouse antibody, Roche, at 1:1000 dilution).

4.11 cDNA preparation and qRT-PCR

For gonad specific qRT-PCR, 20 gonads per replicate were dissected and RNA was isolated using Rneasy Micro Kit (Qiagen) according to the manufacturer's instructions. cDNA was generated using oligo-dT and random hexamer primers with the GoScript Reverse Transcriptase Kit (Promega) was used according to the manufacturer's instructions.

4.12 Western blots

BAT1558 worms were checked for the presence of a possible F55A3.7 protein. BAT748 was used as a positive control for the anti-FLAG-HRP antibody. Those worms were heat-shocked (30 min @ 37°C) and subsequently 42 worms were picked. Washed off N2 worms served as a negative control for the anti-FLAG-HRP antibody. Worms of BAT1749 and BAT1753 were checked for the presence of HMG3::3xHA, SPT-16 and LIN-53 and α-tubulin was used as a loading control. Synchronized young adult worms were washed and pelleted (300 g, 1 minute). Pellets were resuspended in 1x SDS sample buffer (60 mM Tris-HCl (pH 6,8), 10% glycerol, 2% SDS, 1% 14,3 M β-2-mercaptoethanol, 0,02% bromophenol blue, final concentrations). To lyse the samples, they were boiled at 95 °C for 10 minutes buffer. Next, the worm lysates were centrifuged for 10 minutes at 20000 g. Worm lysates were then separated a 10% SDS-PAGE and proteins were transferred onto nitrocellulose blotting membranes (GE Healthcare Life Sciences). Subsequently, the membranes were blocked in Tris-Buffered Saline with 0,1% (w/v) Tween-20 (TBST) containing 0.5% Gelatine Silber (Roth). For protein detection, the blots were probed with primary antibodies (see Table 4) in 3% TBST containing 3% BSA at 4°C. If secondary antibodies were needed, blots were washed in TBST and incubated with secondary antibodies (see Table 4). After a final washing step with TBST, blots were developed using Lumi-Light Western Blotting Substrate (Roche) and analysed with the ImageQuant LAS4000 software (GE Healthcare Life Sciences). To check for α-tubulin, blots were stripped: 2x 10 minutes wash in stripping buffer (15 g Glycine, 1 g SDS, 10 ml Tween20, in 1 l ddH2O; pH 2,2), 2x washed in PBS for 10 minutes and 2x washed in TBST for 5 minutes. Next the blots were blocked and incubated with primary and secondary antibodies as described previously.

Table 4. Antibodies used for Western blots.

Primary antibody	Clonality	Host	Dilution	Company	Secondary antibody
Anti-α-tubulin	Monoclonal	Mouse	1:10000	Sigma	Anti-Mouse-HRP 1:10000 Santa Cruz Biotechnology Inc.
Anti-FLAG-HRP	Monoclonal	Mouse	1:1000	Sigma	-
Anti-HA-HRP	Polyclonal	Chicken	1:1000	Abcam	-
Anti-LIN-53	Monoclonal	Rabbit	1:2000	Pineda	Anti-rabbit-HRP 1:10000 Santa Cruz Biotechnology Inc.
Anti-SPT-16	Monoclonal	Guinea pig	1:2000	Pineda	Anti-Guinea pig-HRP 1:10000 Santa Cruz Biotechnology Inc.

4.13 Transdifferentiation assays (heat-shock induction of *che-1*)

To induce *che-1* expression in *hsp-16.2::che-1* strains, L4 larvae were incubated at 37 °C for 30 minutes and kept overnight at 25 °C.

4.14 RNA-seq

For gonad specific qRT-PCR, 20 gonads per replicate were dissected and RNA was isolated using Rneasy Micro Kit (Qiagen) according to the manufacturer's instructions. Ribosomal RNA depletion (Ribo-Zero, Illumina) was performed according to the manufacturer's instructions. The libraries were prepared using NEXTflex Rapid Directional qRNA-Seq Kit (Bioo Scientific) according to the manufacturer's instructions. The libraries were sequenced using paired end sequencing length of 75 nucleotides on a HiSeq4000 machine (Illumina).

4.15 ATAC-seq

For gonad specific ATAC-seq, 20 gonads per replicate were dissected and nuclei were isolated using a plastic dounce homogenizer with 50 strokes tight-fitting insert in buffer A (15 mM Tris–HCl pH7,5, 2 mM MgCl2, 340 mM sucrose, 0,2 mM spermine, 0,5 mM spermidine, 0,5 mM phenylmethanesulfonate [PMSF], 1mM DTT, 0,1% Trition X-100 and 0.25% NP-40 substitute) as described before[134,135]. The debris was removed by spinning at 100g for 5 minutes and nuclei were counted by Methylene blue staining. 100000 nuclei per sample were pelleted by spinning at 1000g for 10 minutes and proceeded immediately to transposition step

of the ATACseq protocol as described previously[114]. Libraries were amplified for a total of 10 to 18 cycles and sequenced using paired-end-sequencing length of 75 nucleotides using NextSeq 500/550 High Output v2 kit (Illumina).

4.16 List of worm strains used

BAT28	*otIs305 (hsp::che-1::3xHA) V; ntIs1 (gcy-5::GFP) V*
BAT60	*ccIs4251 [myo-3p::DFP(NLS)::LacZ (pSAK2) + myo-3p::GFP(mitochondrially targeted) (pSAK4) + dpy-20 (+)] I. stIs10166 [dpy-7p::HIS-24::mCherry + unc-119(+)] III. barIs19 [hsp-16.2::hlh-1 + rol-6 (su1006]) X.*
BAT173	*barIs19 (hsp-16.2p::hlh-1+rol-6), ccls4251 I [myo-3p::gfp::NLS], stIs10166 (dpy-7p::mCherry), bar18 V.*
BAT197	*ccls4251 I [myo-3p::gfp::NLS], unc-83(bar18) V.*
BAT372	*F55A3.7(ok1829) I.; otIs305 [hsp::che-1::3xHA] ntIs1 [gcy-5::GFP] V.*
BAT653	*him-8(e1489) IV.*
BAT661	*ccls4251 I [myo-3p::gfp::NLS].*
BAT748	*barEx308 [hsp-16.2::H3_S10E::2xFLAG, hsp-16.2::His-71_S10E::2xFLAG, myo-2p::mCherry], otIs305 [hsp-16.2p::che-1::3xHA, rol-6(su1006)] ntIs1 [gcy-5p::GFP, lin-15(+)] V.*
BAT968	*ccls4251 [myo-3p::gfp] I; unc-84(e1174) X.*
BAT1099	*rrrSi261 [myo-3pr::cey-4genomic::operon(gfp::h2b::tbb 3'UTR); unc-119(+)] I. unc-83 (bar18) V.*
BAT1298	*barEx453 [myo-3p::unc-83_cDNA::SL2::NLS::tagRFP::tbb-2_3'UTR, myo-2p::mCherry, HygR].*
BAT1300	*barIs219 [eft-3p::unc-83_cDNA::SL2::NLS::tagRFP::tbb-2_3'UTR, myo-2p::mCherry, HygR]*
BAT1468	*barIs165 [myo-3p::lmn-1T40I::2xFLAG::SL2::NLS::tagRFP]*
BAT1488	*rrrSi261 [myo-3pr::cey-4genomic::operon(gfp::h2b::tbb 3'UTR); unc-119(+)] I.*
BAT1559	*hmg-3(bar24[hmg-3::3xHA]) I*
BAT1594	*F55A3.7-16(bar23[F55A3.7::3xFLAG]) I.*
BAT1749	*hmg-3(bar24[hmg-3::3xHA]) I; F55A3.7(ok1829) I.*
BAT1753	*hmg-3(bar24[hmg-3::3xHA]) I.*
BAT1906	*barEx1031 [myo-3p::3xFLAG::unc-83::unc-54_3′_UTR + ttx-3::mCherry + HygR]; ccls4251 [myo-3p::gfp]; unc-83(bar18) V.*
BAT1907	*barEx1032 [myo-3p::3xFLAG::unc-83::unc-54_3′_UTR + ttx-3::mCherry + HygR]; ccls4251 [myo-3p::gfp]; unc-83(bar18) V.*
BAT1909	*barEx1034 [myo-3p::3xFLAG::unc-83::unc-54_3′_UTR + ttx-3::mCherry + HygR]; ccls4251 [myo-3p::gfp]; unc-83(bar18) V.*
BAT1956	*F55A3.7(ok1829) I.; otIs305 [hsp::che-1::3xHA] ntIs1 [gcy-5::GFP] V; barEx1023 (sspt-16p::sspt-16::sspt-16_UTR, ttx-3::mCherry, HygR)*
BAT1957	*F55A3.7(ok1829) I.; otIs305 [hsp::che-1::3xHA] ntIs1 [gcy-5::GFP] V; barEx1024 (sspt-16p::sspt-16::sspt-16_UTR, ttx-3::mCherry, HygR)*
BAT1959	*F55A3.7(ok1829) I.; otIs305 [hsp::che-1::3xHA] ntIs1 [gcy-5::GFP] V; barEx1026 (sspt-16p::sspt-16::sspt-16_UTR, ttx-3::mCherry, HygR)*
BAT1967	*hmg-4(bar32[hmg-4::3xHA]) III*
N2	*C. elegans* wild isolate. (Bristol)
CB4856	*C. elegans* wild isolate. (Hawaii)
RB1524	*F55A3.7(ok1829) I.*

4.17 List of plasmids used

Addgene #46168	*Peft-3::cas9-SV40_NLS::tbb-2 3'UTR*
pRF4	*rol-6(su1006)*
pCFJ90	*Pmyo-2:mCherry:unc-54 UTR*
dBT618	*dpy-10 sgRNA*
dBT620	*hmg-3 sgRNA*
dBT621	*hmg-4sgRNA*
dBT625	*F55A3.7 sgRNA*

5 References

1. Kolundzic, E. *et al.* FACT Sets a Barrier for Cell Fate Reprogramming in Caenorhabditis elegans and Human Cells. *Dev. Cell* (2018). doi:10.1016/j.devcel.2018.07.006

2. Gurdon, J. B., Elsdale, T. R. & Fischberg, M. Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. *Nature* **182**, 64–65 (1958).

3. Campbell, K. H., McWhir, J., Ritchie, W. A. & Wilmut, I. Sheep cloned by nuclear transfer from a cultured cell line. *Nature* **380**, 64–66 (1996).

4. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987–1000 (1987).

5. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).

6. Sánchez Alvarado, A. & Yamanaka, S. Rethinking differentiation: stem cells, regeneration, and plasticity. *Cell* **157**, 110–119 (2014).

7. Luxton, G. W. G. & Starr, D. A. KASHing up with the nucleus: novel functional roles of KASH proteins at the cytoplasmic surface of the nucleus. *Curr. Opin. Cell Biol.* **28**, 69–75 (2014).

8. Tapley, E. C. & Starr, D. A. Connecting the nucleus to the cytoskeleton by SUN-KASH bridges across the nuclear envelope. *Curr. Opin. Cell Biol.* **25**, 57–62 (2013).

9. Crisp, M. *et al.* Coupling of the nucleus and cytoplasm. *J. Cell Biol.* **172**, 41–53 (2006).

10. Ofenbauer, A. & Tursun, B. The UNC-83/UNC-84 LINC members are required for body wall muscle nuclei positioning in C. elegans. *Matters Sel.* **4**, e201805000009 (2018).

11. Barriere, A. Isolation of C. elegans and related nematodes. *WormBook* (2006). doi:10.1895/wormbook.1.115.1

12. Brenner, S. The genetics of Caenorhabditis elegans. *Genetics* **77**, 71–94 (1974).

13. Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Dev. Biol.* **56**, 110–156 (1977).

14. Sulston, J. E. Neuronal cell lineages in the nematode Caenorhabditis elegans. *Cold Spring Harb. Symp. Quant. Biol.* **48 Pt 2**, 443–452 (1983).

15. Bucher, E. A. & Seydoux, G. Gastrulation in the nematode Caenorhabditis elegans. *Semin. Dev. Biol.* **5**, 121–130 (1994).

16. Hillier, L. W. *et al.* Genomics in C. elegans: so many genes, such a little worm. *Genome Res.* **15**, 1651–1660 (2005).

17. Sonnhammer, E. L. & Durbin, R. Analysis of protein domain families in Caenorhabditis elegans. *Genomics* **46**, 200–216 (1997).

18. Baumeister, R. & Ge, L. The worm in us - Caenorhabditis elegans as a model of human disease. *Trends Biotechnol.* **20**, 147–148 (2002).

19. Barnes, R. & Ruppert, E. *Invertebrate Zoology (Saunders, Philadelphia)*. (1974).

20. van Cleave, H. J. Eutely or Cell Constancy in Its Relation to Body Size. *Q. Rev. Biol.* **7**, 59–67 (1932).

21. Rusin, L. & Malakhov, V. Free living marine nematodes have no eutely. *Dokl. Biol. Sci.* **361**, 132–134 (1998).

22. Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev. Biol.* **100**, 64–119 (1983).

23. Yochem, J. & Herman, R. K. Investigating C. elegans development through mosaic analysis. *Development* **130**, 4761–4768 (2003).

24. Shemer, G. & Podbilewicz, B. Fusomorphogenesis: Cell fusion in organ formation. *Dev. Dyn.* **218**, 30–51 (2000).

25. Hermaphrodite - Muscle System - Introduction. Available at: http://www.wormatlas.org/hermaphrodite/muscleintro/MusIntroframeset.html. (Accessed: 14th April 2018)

26. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode Caenorhabditis elegans. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **314**, 1–340 (1986).

27. MacLeod, A. R., Karn, J. & Brenner, S. Molecular analysis of the unc-54 myosin heavy-chain gene of Caenorhabditis elegans. *Nature* **291**, 386–390 (1981).

28. Miller, D. M., Stockdale, F. E. & Karn, J. Immunological identification of the genes encoding the four myosin heavy chain isoforms of Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 2305–2309 (1986).

29. Immunochemical localization of myosin heavy chain isoforms and paramyosin in developmentally and structurally diverse muscle cell types of the nematode Caenorhabditis elegans. *J. Cell Biol.* **105**, 2763–2770 (1987).

30. Hubbard, E. J. Introduction to the germ line. *WormBook* (2005). doi:10.1895/wormbook.1.18.1

31. Strome, S. Specification of the germ line. *WormBook* (2005). doi:10.1895/wormbook.1.9.1

32. Luria, S. E. & Delbrück, M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* **28**, 491–511 (1943).

33. Barstead, R. J. & Moerman, D. G. C. elegans deletion mutant screening. *Methods Mol. Biol. Clifton NJ* **351**, 51–58 (2006).

34. De Stasio, E. A. & Dorman, S. Optimization of ENU mutagenesis of Caenorhabditis elegans. *Mutat. Res.* **495**, 81–88 (2001).

35. Johnsen, R. C. & Baillie, D. L. Formaldehyde mutagenesis of the eT1 balanced region in Caenorhabditis elegans: dose-response curve and the analysis of mutational events. *Mutat. Res.* **201**, 137–147 (1988).

36. Greenwald, I. S. & Horvitz, H. R. unc-93(e1500): A behavioral mutant of Caenorhabditis elegans that defines a gene with a wild-type null phenotype. *Genetics* **96**, 147–164 (1980).

37. Anderson, P. & Brenner, S. A selection for myosin heavy chain mutants in the nematode Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 4470–4474 (1984).

38. Trent, C. *et al.* Sex-specific transcriptional regulation of the C. elegans sex-determining gene her-1. *Mech. Dev.* **34**, 43–55 (1991).

39. Brookes, P. & Lawley, P. D. The reaction of mono- and di-functional alkylating agents with nucleic acids. *Biochem. J.* **80**, 496–503 (1961).

40. Coulondre, C. & Miller, J. H. Genetic studies of the lac repressor. III. Additional correlation of mutational sites with specific amino acid residues. *J. Mol. Biol.* **117**, 525–567 (1977).

41. Flibotte, S. *et al.* Whole-genome profiling of mutagenesis in Caenorhabditis elegans. *Genetics* **185**, 431–441 (2010).

42. Anderson, P. Mutagenesis. *Methods Cell Biol.* **48**, 31–58 (1995).

43. Jorgensen, E. M. & Mango, S. E. The art and design of genetic screens: caenorhabditis elegans. *Nat. Rev. Genet.* **3**, 356–369 (2002).

44. Stewart, H. I., Rosenbluth, R. E. & Baillie, D. L. Most ultraviolet irradiation induced mutations in the nematode Caenorhabditis elegans are chromosomal rearrangements. *Mutat. Res.* **249**, 37–54 (1991).

45. Rosenbluth, R. E., Cuddeford, C. & Baillie, D. L. Mutagenesis in Caenorhabditis elegans. II. A spectrum of mutational events induced with 1500 r of gamma-radiation. *Genetics* **109**, 493–511 (1985).

46. Babu, P. & Brenner, S. Spectrum of 32P-induced mutants of Caenorhabditis elegans. *Mutat. Res.* **82**, 269–273 (1981).

47. Smith, H. E. Identifying insertion mutations by whole-genome sequencing. *BioTechniques* **50**, 96–97 (2011).

48. Williams, D. C., Boulin, T., Ruaud, A.-F., Jorgensen, E. M. & Bessereau, J.-L. Characterization of Mos1-mediated mutagenesis in Caenorhabditis elegans: a method for the rapid identification of mutated genes. *Genetics* **169**, 1779–1785 (2005).

49. Bessereau, J. L. *et al.* Mobilization of a Drosophila transposon in the Caenorhabditis elegans germ line. *Nature* **413**, 70–74 (2001).

50. Robert, V. & Bessereau, J.-L. Targeted engineering of the Caenorhabditis elegans genome following Mos1-triggered chromosomal breaks. *EMBO J.* **26**, 170–183 (2007).

51. Frøkjaer-Jensen, C. *et al.* Targeted gene deletions in C. elegans using transposon excision. *Nat. Methods* **7**, 451–453 (2010).

52. Frøkjær-Jensen, C., Davis, M. W., Ailion, M. & Jorgensen, E. M. Improved Mos1-mediated transgenesis in C. elegans. *Nat. Methods* **9**, 117–118 (2012).

53. Frøkjaer-Jensen, C. *et al.* Single-copy insertion of transgenes in Caenorhabditis elegans. *Nat. Genet.* **40**, 1375–1383 (2008).

54. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1156–1160 (1996).

55. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**, 143–148 (2011).

56. Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331–338 (2012).

57. Ott de Bruin, L. M., Volpi, S. & Musunuru, K. Novel Genome-Editing Tools to Model and Correct Primary Immunodeficiencies. *Front. Immunol.* **6**, 250 (2015).

58. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).

59. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).

60. Guilinger, J. P., Thompson, D. B. & Liu, D. R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **32**, 577–582 (2014).

61. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806–811 (1998).

62. Chapman, E. J. & Carrington, J. C. Specialization and evolution of endogenous small RNA pathways. *Nat. Rev. Genet.* **8**, 884–896 (2007).

63. Boutros, M. & Ahringer, J. The art and design of genetic screens: RNA interference. *Nat. Rev. Genet.* **9**, 554 (2008).

64. Sijen, T., Steiner, F. A., Thijssen, K. L. & Plasterk, R. H. A. Secondary siRNAs Result from Unprimed RNA Synthesis and Form a Distinct Class. *Science* **315**, 244–247 (2007).

65. Pak, J. & Fire, A. Distinct Populations of Primary and Secondary Effectors During RNAi in C. elegans. *Science* **315**, 241–244 (2007).

66. Aoki, K., Moriguchi, H., Yoshioka, T., Okawa, K. & Tabara, H. In vitro analyses of the production and activity of secondary small interfering RNAs in C. elegans. *EMBO J.* **26**, 5007–5019 (2007).

67. Zhang, C. *et al.* The Caenorhabditis elegans RDE-10/RDE-11 Complex Regulates RNAi by Promoting Secondary siRNA Amplification. *Curr. Biol.* **22**, 881–890 (2012).

68. Ahringer, J. Reverse genetics. *WormBook* (2006). doi:10.1895/wormbook.1.47.1

69. Timmons, L., Court, D. L. & Fire, A. Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in Caenorhabditis elegans. *Gene* **263**, 103–112 (2001).

70. Fraser, A. G. *et al.* Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. *Nature* **408**, 325–330 (2000).

71. Kamath, R. S. *et al.* Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* **421**, 231–237 (2003).

72. C. elegans RNAi Collection (Ahringer) | Source BioScience. Available at: https://www.sourcebioscience.com/products/life-sciences-research/clones/rnai-resources/c-elegans-rnai-collection-ahringer/. (Accessed: 12th July 2018)

73. Rual, J.-F. *et al.* Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library. *Genome Res.* **14**, 2162–2168 (2004).

74. C. elegans ORF-RNAi Resource (Vidal) | Source BioScience. Available at: https://www.sourcebioscience.com/products/life-sciences-research/clones/rnai-resources/c-elegans-orf-rnai-resource-vidal/. (Accessed: 12th July 2018)

75. Horvitz, H. R. & Sulston, J. E. Isolation and Genetic Characterization of Cell-Lineage Mutants of the Nematode Caenorhabditis Elegans. *Genetics* **96**, 435–454 (1980).

76. Malone, C. J., Fixsen, W. D., Horvitz, H. R. & Han, M. UNC-84 localizes to the nuclear envelope and is required for nuclear migration and anchoring during C. elegans development. *Dev. Camb. Engl.* **126**, 3171–3181 (1999).

77. Sulston, J. E. & Horvitz, H. R. Abnormal cell lineages in mutants of the nematode Caenorhabditis elegans. *Dev. Biol.* **82**, 41–55 (1981).

78. McGee, M. D., Rillo, R., Anderson, A. S. & Starr, D. A. UNC-83 IS a KASH protein required for nuclear migration and is recruited to the outer nuclear membrane by a physical interaction with the SUN protein UNC-84. *Mol. Biol. Cell* **17**, 1790–1801 (2006).

79. Fridolfsson, H. N., Ly, N., Meyerzon, M. & Starr, D. A. UNC-83 coordinates kinesin-1 and dynein activities at the nuclear envelope during nuclear migration. *Dev. Biol.* **338**, 237–250 (2010).

80. Fridolfsson, H. N. & Starr, D. A. Kinesin-1 and dynein at the nuclear envelope mediate the bidirectional migrations of nuclei. *J. Cell Biol.* **191**, 115–128 (2010).

81. Starr, D. A. Watching nuclei move: Insights into how kinesin-1 and dynein function together. *Bioarchitecture* **1**, 9–13 (2011).

82. Bone, C. R., Chang, Y.-T., Cain, N. E., Murphy, S. P. & Starr, D. A. Nuclei migrate through constricted spaces using microtubule motors and actin networks in C. elegans hypodermal cells. *Dev. Camb. Engl.* **143**, 4193–4202 (2016).

83. Tursun, B., Patel, T., Kratsios, P. & Hobert, O. Direct conversion of C. elegans germ cells into specific neuron types. *Science* **331**, 304–308 (2011).

84. Patel, T., Tursun, B., Rahe, D. P. & Hobert, O. Removal of Polycomb repressive complex 2 makes C. elegans germ cells susceptible to direct conversion into specific somatic cell types. *Cell Rep.* **2**, 1178–1186 (2012).

85. Cheloufi, S. *et al.* The histone chaperone CAF-1 safeguards somatic cell identity. *Nature* **528**, 218–224 (2015).

86. Orphanides, G., LeRoy, G., Chang, C. H., Luse, D. S. & Reinberg, D. FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* **92**, 105–116 (1998).

87. Hondele, M. *et al.* Structural basis of histone H2A-H2B recognition by the essential chaperone FACT. *Nature* **499**, 111–114 (2013).

88. Keller, D. M. & Lu, H. p53 serine 392 phosphorylation increases after UV through induction of the assembly of the CK2.hSPT16.SSRP1 complex. *J. Biol. Chem.* **277**, 50206–50213 (2002).

89. Wittmeyer, J. & Formosa, T. The Saccharomyces cerevisiae DNA polymerase alpha catalytic subunit interacts with Cdc68/Spt16 and with Pob3, a protein similar to an HMG1-like protein. *Mol. Cell. Biol.* **17**, 4178–4190 (1997).

90. Schlesinger, M. B. & Formosa, T. POB3 is required for both transcription and replication in the yeast Saccharomyces cerevisiae. *Genetics* **155**, 1593–1606 (2000).

91. Yang, J. *et al.* The Histone Chaperone FACT Contributes to DNA Replication-Coupled Nucleosome Assembly. *Cell Rep.* **14**, 1128–1141 (2016).

92. Kaplan, C. D., Laprade, L. & Winston, F. Transcription elongation factors repress transcription initiation from cryptic sites. *Science* **301**, 1096–1099 (2003).

93. Becker, J. S., Nicetto, D. & Zaret, K. S. H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet.* **32**, 29–41 (2016).

94. Vanin, E. F. Processed pseudogenes: characteristics and evolution. *Annu. Rev. Genet.* **19**, 253–272 (1985).

95. Chandrasekaran , C. & Betrán , E. Origins of New Genes and Pseudogenes. *Nat. Educ.* **1**, 181 (2008).

96. Xiao-Jie, L., Ai-Mei, G., Li-Juan, J. & Jiang, X. Pseudogene in cancer: real functions and promising signature. *J. Med. Genet.* **52**, 17–24 (2015).

97. Jurka, J. Evolutionary impact of human Alu repetitive elements. *Curr. Opin. Genet. Dev.* **14**, 603–608 (2004).

98. Dewannieux, M. & Heidmann, T. LINEs, SINEs and processed pseudogenes: parasitic strategies for genome modeling. *Cytogenet. Genome Res.* **110**, 35–48 (2005).

99. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155 (2000).

100. Nishikimi, M., Fukuyama, R., Minoshima, S., Shimizu, N. & Yagi, K. Cloning and chromosomal mapping of the human nonfunctional gene for L-gulono-gamma-lactone oxidase, the enzyme for L-ascorbic acid biosynthesis missing in man. *J. Biol. Chem.* **269**, 13685–13688 (1994).

101. Prieto-Godino, L. L. *et al.* Olfactory receptor pseudo-pseudogenes. *Nature* **539**, 93–97 (2016).

102. Pei, B. *et al.* The GENCODE pseudogene resource. *Genome Biol.* **13**, R51 (2012).

103. Poliseno, L., Marranci, A. & Pandolfi, P. P. Pseudogenes in Human Cancer. *Front. Med.* **2**, 68 (2015).

104. Ena Kolundzic. Identification and characterization of factors that regulate cell plasticity in C. elegans. (Humboldt-Universität zu Berlin, 2017).

105. Doitsidou, M., Poole, R. J., Sarin, S., Bigelow, H. & Hobert, O. C. elegans Mutant Identification with a One-Step Whole-Genome-Sequencing and SNP Mapping Strategy. *PLoS ONE* **5**, (2010).

106. Minevich, G., Park, D. S., Blankenberg, D., Poole, R. J. & Hobert, O. CloudMap: a cloud-based pipeline for analysis of mutant genome sequences. *Genetics* **192**, 1249–1269 (2012).

107. Flavell, S. W. *et al.* Serotonin and the Neuropeptide PDF Initiate and Extend Opposing Behavioral States in C. elegans. *Cell* **154**, 1023–1035 (2013).

108. Nussbaum-Krammer, C. I., Neto, M. F., Brielmann, R. M., Pedersen, J. S. & Morimoto, R. I. Investigating the spreading and toxicity of prion-like proteins using the metazoan model organism C. elegans. *J. Vis. Exp. JoVE* 52321 (2015). doi:10.3791/52321

109. Han, S. K. *et al.* OASIS 2: online application for survival analysis 2 with features for the analysis of maximal lifespan and healthspan in aging research. *Oncotarget* **7**, 56147–56152 (2016).

110. Dorman, J. B., Albinder, B., Shroyer, T. & Kenyon, C. The age-1 and daf-2 genes function in a common pathway to control the lifespan of Caenorhabditis elegans. *Genetics* **141**, 1399–1406 (1995).

111. Lapierre, L. R. & Hansen, M. Lessons from C. elegans: signaling pathways for longevity. *Trends Endocrinol. Metab. TEM* **23**, 637–644 (2012).

112. Nandi, D., Tahiliani, P., Kumar, A. & Chandu, D. The ubiquitin-proteasome system. *J. Biosci.* **31**, 137–155 (2006).

113. Beck, F. *et al.* Near-atomic resolution structural model of the yeast 26S proteasome. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 14870–14875 (2012).

114. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

115. Buenrostro, J., Wu, B., Chang, H. & Greenleaf, W. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol. Ed. Frederick M Ausubel Al* **109**, 21.29.1 (2015).

116. eef-1A.1 (gene) - WormBase : Nematode Information Resource. Available at: https://wormbase.org/species/c_elegans/gene/WBGene00001168#0-9g-3. (Accessed: 24th October 2018)

117. col-19 (gene) - WormBase : Nematode Information Resource. Available at: https://www.wormbase.org/species/c_elegans/gene/WBGene00000608#0-9g-3. (Accessed: 24th October 2018)

118. vit-5 (gene) - WormBase : Nematode Information Resource. Available at: https://www.wormbase.org/species/c_elegans/gene/WBGene00006929#0-9g-3. (Accessed: 24th October 2018)

119. Hubbard, E. J. A. FLP/FRT and Cre/lox recombination technology in C. elegans. *Methods San Diego Calif* **68**, 417–424 (2014).

120. Wang, H., Liu, J., Yuet, K. P., Hill, A. J. & Sternberg, P. W. Split cGAL, an intersectional strategy using a split intein for refined spatiotemporal transgene control in Caenorhabditis elegans. *Proc. Natl. Acad. Sci.* **115**, 3900–3905 (2018).

121. Uzer, G., Rubin, C. T. & Rubin, J. Cell Mechanosensitivity is Enabled by the LINC Nuclear Complex. *Curr. Mol. Biol. Rep.* **2**, 36–47 (2016).

122. Isermann, P. & Lammerding, J. Nuclear Mechanics and Mechanotransduction in Health and Disease. *Curr. Biol.* **23**, R1113–R1121 (2013).

123. Mello, C. & Fire, A. Chapter 19 DNA Transformation. in *Methods in Cell Biology* (eds. Epstein, H. F. & Shakes, D. C.) **48**, 451–482 (Academic Press, 1995).

124. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).

125. Dibb, N. J., Maruyama, I. N., Krause, M. & Karn, J. Sequence analysis of the complete Caenorhabditis elegans myosin heavy chain gene family. *J. Mol. Biol.* **205**, 603–613 (1989).

126. Radman, I., Greiss, S. & Chin, J. W. Efficient and Rapid C. elegans Transgenesis by Bombardment and Hygromycin B Selection. *PLoS ONE* **8**, (2013).

127. Robust expression of transgenes in the C. elegans germline through a simple microinjection protocol. *The WBG* (2013). Available at: http://wbg.wormbook.org/2013/02/21/robust-expression-of-transgenes-in-the-c-elegans-germline-through-a-simple-microinjection-protocol/. (Accessed: 7th November 2018)

128. Kelly, W. G., Xu, S., Montgomery, M. K. & Fire, A. Distinct requirements for somatic and germline expression of a generally expressed Caernorhabditis elegans gene. *Genetics* **146**, 227–238 (1997).

129. Arribere, J. A. *et al.* Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans. *Genetics* **198**, 837–846 (2014).

130. Agarose immobilization of C. elegans. *The WBG* (2009). Available at: http://wbg.wormbook.org/2009/12/01/agarose-immobilization-of-c-elegans/. (Accessed: 13th April 2018)

131. Hosono, R. Sterilization and growth inhibition of Caenorhabditis elegans by 5-fluorodeoxyuridine. *Exp. Gerontol.* **13**, 369–374 (1978).

132. Amrit, F. R. G., Ratnappan, R., Keith, S. A. & Ghazi, A. The C. elegans lifespan assay toolkit. *Methods San Diego Calif* **68**, 465–475 (2014).

133. Bettinger, J. C., Lee, K. & Rougvie, A. E. Stage-specific accumulation of the terminal differentiation factor LIN-29 during Caenorhabditis elegans development. *Dev. Camb. Engl.* **122**, 2517–2527 (1996).

134. Ooi, S. L., Henikoff, J. G. & Henikoff, S. A native chromatin purification system for epigenomic profiling in Caenorhabditis elegans. *Nucleic Acids Res.* **38**, e26 (2010).

135. Steiner, F. A., Talbert, P. B., Kasinathan, S., Deal, R. B. & Henikoff, S. Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. *Genome Res.* **22**, 766–777 (2012).

Publications

Parts of this thesis were published in the following publications:

Ofenbauer A. & Tursun B. The UNC-83/UNC-84 LINC members are required for body wall muscle nuclei positioning in C. elegans. *Matters Sel.*, (2018).

Kolundzic E., **Ofenbauer A.**, Bulut SI., Uyar B., Baytek G., Sommermeier A., Seelk S., He M., Hirsekorn A., Vucicevic D., Akalin A., Diecke S., Lacadie SA., Tursun B. FACT Sets a Barrier for Cell Fate Reprogramming in Caenorhabditis elegans and Human Cells. *Dev. Cell* (2018).

Other publications published during my time as a PhD student, that aren't part of this thesis:

Hajduskova M., Baytek G., Kolundzic E., Gosdschan A., Kazmierczak M., **Ofenbauer A.**, Beato Del Rosal ML., Herzog S., Ul Fatima N., Mertins P., Seelk-Müthel S., Tursun B. MRG-1/MRG15 Is a Barrier for Germ Cell to Neuron Reprogramming in Caenorhabditis elegans. *Genetics* (2018).

Ofenbauer, A., Sebinger, D. D. R., Prewitz, M., Gruber, P. & Werner, C. Dewaxed ECM: A simple method for analyzing cell behaviour on decellularized extracellular matrices. *J. Tissue Eng. Regen. Med.* **9**, 1046–1055 (2015).

Sebinger, D. D. R., **Ofenbauer, A.**, Gruber, P., Malik, S. & Werner, C. ECM modulated early kidney development in embryonic organ culture. *Biomaterials* **34**, 6670–6682 (2013).

Acknowledgements

I would like to thank a lot of people that supported me in various ways to make this thesis happen.

First of all I would like to thank Dr. Baris Tursun, not only for giving me the opportunity to work in his lab and for taking me to fancy conferences, but also for being a kind mentor and continuously supporting me. I am very glad that what I could learn from him was also going far beyond just hard science facts.

I also would like to thank the members of my PhD-committee, Prof. Dr. Andreas Herrmann, Prof. Dr. Thomas Sommer, Dr. Baris Tursun, Prof. Dr. Andrew Plested and Prof. Dr. Uwe Ohler.

I would like to thank all my co-workers for the warm working atmosphere of our wormy lab, the scientific discussions and the fun we had every day. Thanks for all the hugs and support. In particular I would like thank: Ena, Steffi and Martina, who were there from the beginning, for being open and welcoming to connect also outside the lab. Selman and Gülkiz for sharing “deli” stuff with me and for still having the power and courage to be friends with me. Anna for being one of the kindest person I know of and especially for being so kind as to read through my thesis and giving me helpful suggestions.

For experimental help I would like to thank my students Anne Sommermeier and Kitty van Scharenburg, it was a wonderful experience working with you.

Finally I want to thank my family, my friends and David for all the love and the wonderful connections we’ve build - without you I wouldn’t be me.

Declaration

Erklärung über die selbstständige Abfassung meiner Dissertation

Hiermit erkläre ich, Andreas Ofenbauer, Matrikel-Nr: 519312, dass ich die vorliegende Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Dissertation wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt oder veröffentlicht.

Berlin, am 22.11.2018

Unterschrift